

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 24, 2004, 06:40:16 ; Search time 523 Seconds
(without alignments)
2371.843 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKTIIDGKLVAKQIREBIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPFC.spool.p/US10723061/runat.21072004.090400.7098/app.query.fasta_1.455
-DB=N_Geneseq_29Jan04 -QWMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNIT5=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061 @CGN 1 1 885 @runat.21072004.090400.7098 -NCPU=6 -ICPU=3
-NO.WMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=130 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1486	100.0	1116	7 ADA70516	Ada70516 Rice gene
2	1238	83.3	1083	6 ABZ12736	Abz12736 Arabidops
3	1238	83.3	1083	7 ADA68129	Ada68129 Arabidops
4	1238	83.3	1282	3 AAC42303	Aac42303 Arabidops
5	1035	69.7	1086	3 AAC34458	Aac34458 Arabidops
6	1035	69.7	1182	3 AAC39538	Aac39538 Arabidops
7	1017	68.4	1688	3 AAC45046	Aac45046 Arabidops
8	1012	68.1	1327	3 AAC36669	Aac36669 Arabidops

9	949	63.9	1269	7 ADA71098	Ada71098 Rice gene
10	763.5	51.4	110000	6 ABQ69245_13	Continuation (14 o
11	763.5	51.4	110000	6 ABQ67195_3	Continuation (4 of
12	762	51.3	3003	4 ABL07155	Ab107155 Drosophil
13	755.5	50.8	110000	6 ABA03041_13	Continuation (14 o
14	751.5	50.6	5635	6 ABQ71010	Abq71010 Listeria
15	743.5	50.0	858	6 ABN66178	Abn66178 Streptoco
16	739.5	49.8	2000	7 ACC60718	Acc60718 Gene sequ
17	739.5	49.8	2841	7 ABZ70970	Abz70970 Saccharom
18	735.5	49.5	4474	9 ADB76850	Adb76850 A. gossyp
19	735	49.5	3112	5 AAC91213	Aac91213 Human MTH
20	735	49.5	3112	6 ABL68500	Ab168500 Kidney ca
21	735	49.5	3112	6 ABN95605	Abn95605 Gene #210
22	735	49.5	3179	5 Aas66896	Aas66896 DNA encod
23	731	49.2	3332	9 ADB57908	Adb57908 Toxicity-
24	724	48.7	855	7 ABX06500	Abx06500 S. pneumo
25	724	48.7	858	3 AAA05466	Aaa05466 Streptoco
26	724	48.7	13440	2 AAV52261	Aav52261 Streptoco
27	724	48.7	110000	7 ABS56454_06	Continuation (7 of
28	721	48.5	852	6 ABN66177	Abn66177 Streptoco
29	721	48.5	110000	6 ABN71527_04	Continuation (5 of
30	713.5	48.0	2000	7 ACC60721	Acc60721 Gene sequ
31	712.5	47.9	855	9 ADC94007	Adc94007 B. faecali
32	708.5	47.7	110000	6 ABA90521_08	Continuation (9 of
33	704.5	47.4	7528	2 AAX12992	Aax12992 Enterococ
34	704.5	47.4	7528	6 ABS98787	Abs98787 Enterococ
35	687.5	46.3	855	6 ABK73707	Abk73707 Bacillus
36	687	46.2	110000	3 AAA81489_7	Continuation (8 of
37	687	46.2	172325	3 AAF21613	Aaf21613 Neisseria
38	681.5	45.9	110000	2 AAT42063_06	Continuation (7 of
39	679.5	45.7	873	6 ABN91101	Abn91101 Staphyloc
40	678	45.6	786	7 ABZ37976	Abz37976 N. gonorr
41	675	45.4	876	7 ACF71018	Acf71018 Photorhab
42	675	45.4	110000	7 AC567367_41	Continuation (42 o
43	675	45.4	110000	7 AC565388_06	Continuation (7 of
44	670.5	45.1	882	7 ACF74867	Acf74867 Staphyloc
45	670.5	45.1	9062	2 AAV74396	Aav74396 Staphyloc

ALIGNMENTS

RESULT 1

ADA70516
ID ADA70516 standard; DNA; 1116 BP.

AC ADA70516;

DT 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 3839.

XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX WPI; 2003-175290/17.

XX Identifying at least one gene involved in plant resistance or response to

XX pathogenic infection for conferring resistance or tolerance to a plant to

PT


```
US-10-723-061-18 (1-252) x ABZ12736 (1-1083)
QY 1 AlalysileAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 205 GCAATTGTAATGATGGAAGCGTGTGGCAAAAAGATTAGAGTGAATCAACAATTGAA 264
QY 21 IleAlalysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 265 GTTTCAGAAATCAAGGAATCAATGCTGTGATCTCTGTTAGCAGTAATCTTTGGT 324
QY 41 SerArgLysAspSerGlnThrTyValArgAsnLysLysAlaCysGluAlaValGly 60
Db 325 GACAGAAAGATTCTCGCACTTATGTGAGGAACAAGAAAGCTGTGACCTCGTTGA 384
QY 61 IleLysSerTyGluValAsnLeuProGluAAspSerSerGluAspGluValLeuLysHis 80
Db 385 ATCAAAATCGTTCGAAGTTCGTCTAGCTGAAGATTTCATCAGAAGAGAGGTGTGAAATCT 444
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGGATTCATATGATGATCTTCTGTCATGATCTTCTGTCAGTTCCTCTGCCA 504
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGGATGACAGAAACATATCGATCGGTTAGTATAGAGAAGATGTTGACGGA 564
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 565 TTTTCATCGCTAAATATATGACGGCTTCCCATGCGTGGAGAGAACCTTATTCGTCTCT 624
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyGlyValGluIleLysGlyLys 160
Db 625 TGTACTCCAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
QY 161 ArgAlaValAlaIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGAGCGGTGTTATCGGAAGAGTAACATTCGCGTATGCCAGCTGCTTTTACTCGAG 744
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 745 AGGAGGATGCAACCGTTAGCATATTCATTCAGAACCAAGAACCCCTGGAAGAAATCACA 804
QY 201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTyr 220
Db 805 AGAGAAGCTGATATTAATCTCAGCTGTGACAGCCCAACATGTCAGAGGAGCTGG 864
QY 221 IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 865 ATAAACCGCGCGCAGTCTCATCGATGTTGGATTAATCCTGTGAGGATCCCAAGTCT 924
QY 241 ProArgGlyTyArgLeuValGlyAspValCysTyValGluAlaLaserLysIleAlaGly 260
Db 925 GCGCGTGATATCGATTGTTGGACATTTGCTACGAGGAGGCTAGCAAGTTGCATCA 984
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 985 GCATCACACCTGTTCTCTGGCGGTGTAGGACCAATGACCATAGCATGCTTCTATCCAAC 1044
QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1045 ACTTTTAAACATCAGCTAAGAGGATTACCAACTTCCAG 1080
RESULT 3
ID ADA68129 standard; DNA; 1083 BP.
XX
AC ADA68129;
XX
XX 20-NOV-2003 (first entry)
DT
DE Arabidopsis thaliana gene, SEQ ID 393.
DE
XX
```

Plant; bacterial infection; fungal infection; viral infection; ds.
Arabidopsis thaliana.
WO2003000898-A1.
03-JAN-2003.
22-JUN-2001; 2001WO-IB001105.
22-JUN-2001; 2001WO-IB001105.
(SYGN) SYNGENTA PARTICIPATIONS AG.
Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
WPI; 2003-175290/17.
Identifying at least one gene involved in plant resistance or response to
pathogenic infection for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection by determining or detecting plant
gene expression.
Claim 6; SEQ ID NO 393; 899pp; English.
The present invention relates to a method (M1) for identifying genes
involved in plant resistance or response to pathogenic infection. M1
comprises identifying a gene whose expression is significantly altered in
the incompatible interaction of plant gene expression relative to
the expression of the gene in an uninfected plant, in a mutant plant that
does not express a gene associated with response to pathogenic infection,
or in a corresponding incompatible or compatible interaction. (M1) is
useful for conferring resistance to resistance or tolerance to a plant to
bacterial, fungal or viral infection. The present sequence was used to
illustrate the invention.
SQ Sequence 1083 BP; 305 A; 230 C; 259 G; 289 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3,22e-127 Length: 1083
Score: 1238.00 Matches: 231
Percent Similarity: 91.44% Conservative: 36
Best Local Similarity: 79.11% Mismatches: 25
Query Match: 83.31% Indels: 0
DB: 7 Gaps: 0

US-10-723-061-18 (1-292) x ADA68129 (1-1083)
QY 1 AlalysileAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 205 GCAATTGTAATGATGGAAGCGTGTGGCAAAAAGATTAGAGTGAATCAACAATTGAA 264
QY 21 IleAlalysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 265 GTTTCAGAAATCAAGGAATCAATGCTGTGATCTCTGTTAGCAGTAATCTTTGGT 324
QY 41 SerArgLysAspSerGlnThrTyValArgAsnLysLysAlaCysGluAlaValGly 60
Db 325 GACAGAAAGATTCTCGCACTTATGTGAGGAACAAGAAAGCTGTGACCTCGTTGA 384
QY 61 IleLysSerTyGluValAsnLeuProGluAAspSerSerGluAspGluValLeuLysHis 80
Db 385 ATCAAAATCGTTCGAAGTTCGTCTAGCTGAAGATTTCATCAGAAGAGAGGTGTGAAATCT 444
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGGATTCATATGATGATCTTCTGTCATGATCTTCTGTCAGTTCCTCTGCCA 504
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGGATGACAGAAACATATCGATCGGTTAGTATAGAGAAGATGTTGACGGA 564

QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 565 TTTTCATCCGCTAAATATGGACGGCTGGCATGCGTGAAGAGAACCTTATTCGTTCT 624
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 625 TGTACTCCAAAGATGCATGTGATGTTGTCATAGATACACATTCGAATCAAGGAAG 684
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGAGCGGTTCTTATCGAAGAGTAACTGTGCGTATGCCAGTCCTCTTTTACTGCAG 744
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 745 AGGAGGATCAACCGTTAGCATTTCCATTCCTCAAGAACCAAGACCTGAAGAATCA 804
QY 201 ArgGlnAlaAspIleValIleAlaValAlaGlyValAlaAsnLeuValArgGlySerTrp 220
Db 805 AGAAGAGCTGATATTAATCTCAGCTGTGGACGCCAAACATGTCAGAGGAAGCTGG 864
QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 865 ATAAACCGGGCGAGTCCCTCATCGATGTGGATTAACTCCTGTGAGATCCAAAGTGTCT 924
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 925 GCGGTGGATATCGATTGGTTGGAGACATTTGCTACGAGGAGGCTAGCAAGTTGCATCA 984
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaValMetLeuSerAsn 280
Db 985 GCATCACACCTGTTCTCTGCGGTGAGGACCAATGACCATGCCATGCTTCTATCCAAC 1044
QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1045 ACITTAACATCAGCTAGAGGATTCACACTTCCAG 1080
RESULT 4
AAC42303
ID AAC42303 standard; DNA; 1282 BP.
XX
AC AAC42303;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 35038.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 03-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 23-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139464P.
PR 18-JUN-1999; 99US-0139465P.
PR 18-JUN-1999; 99US-0139466P.
PR 18-JUN-1999; 99US-0139467P.
PR 18-JUN-1999; 99US-0139468P.
PR 18-JUN-1999; 99US-0139469P.
PR 18-JUN-1999; 99US-0139470P.
PR 18-JUN-1999; 99US-0139471P.
PR 18-JUN-1999; 99US-0139472P.
PR 18-JUN-1999; 99US-0139473P.
PR 18-JUN-1999; 99US-0139474P.
PR 18-JUN-1999; 99US-0139475P.
PR 18-JUN-1999; 99US-0139476P.
PR 18-JUN-1999; 99US-0139477P.
PR 18-JUN-1999; 99US-0139478P.
PR 18-JUN-1999; 99US-0139479P.
PR 18-JUN-1999; 99US-0139480P.
PR 18-JUN-1999; 99US-0139481P.
PR 18-JUN-1999; 99US-0139482P.
PR 18-JUN-1999; 99US-0139483P.
PR 18-JUN-1999; 99US-0139484P.
PR 18-JUN-1999; 99US-0139485P.
PR 18-JUN-1999; 99US-0139486P.
PR 18-JUN-1999; 99US-0139487P.
PR 18-JUN-1999; 99US-0139488P.
PR 18-JUN-1999; 99US-0139489P.
PR 18-JUN-1999; 99US-0139490P.
PR 18-JUN-1999; 99US-0139491P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139493P.
PR 18-JUN-1999; 99US-0139494P.
PR 18-JUN-1999; 99US-0139495P.
PR 18-JUN-1999; 99US-0139496P.
PR 18-JUN-1999; 99US-0139497P.
PR 18-JUN-1999; 99US-0139498P.
PR 18-JUN-1999; 99US-0139499P.
PR 18-JUN-1999; 99US-0139500P.
PR 18-JUN-1999; 99US-0139501P.
PR 18-JUN-1999; 99US-0139502P.
PR 18-JUN-1999; 99US-0139503P.
PR 18-JUN-1999; 99US-0139504P.
PR 18-JUN-1999; 99US-0139505P.
PR 18-JUN-1999; 99US-0139506P.
PR 18-JUN-1999; 99US-0139507P.
PR 18-JUN-1999; 99US-0139508P.
PR 18-JUN-1999; 99US-0139509P.
PR 18-JUN-1999; 99US-0139510P.
PR 18-JUN-1999; 99US-0139511P.
PR 18-JUN-1999; 99US-0139512P.
PR 18-JUN-1999; 99US-0139513P.
PR 18-JUN-1999; 99US-0139514P.
PR 18-JUN-1999; 99US-0139515P.
PR 18-JUN-1999; 99US-0139516P.
PR 18-JUN-1999; 99US-0139517P.
PR 18-JUN-1999; 99US-0139518P.
PR 18-JUN-1999; 99US-0139519P.
PR 18-JUN-1999; 99US-0139520P.
PR 18-JUN-1999; 99US-0139521P.
PR 18-JUN-1999; 99US-0139522P.
PR 18-JUN-1999; 99US-0139523P.
PR 18-JUN-1999; 99US-0139524P.
PR 18-JUN-1999; 99US-0139525P.
PR 18-JUN-1999; 99US-0139526P.
PR 18-JUN-1999; 99US-0139527P.
PR 18-JUN-1999; 99US-0139528P.
PR 18-JUN-1999; 99US-0139529P.
PR 18-JUN-1999; 99US-0139530P.
PR 18-JUN-1999; 99US-0139531P.
PR 18-JUN-1999; 99US-0139532P.
PR 18-JUN-1999; 99US-0139533P.
PR 18-JUN-1999; 99US-0139534P.
PR 18-JUN-1999; 99US-0139535P.
PR 18-JUN-1999; 99US-0139536P.
PR 18-JUN-1999; 99US-0139537P.
PR 18-JUN-1999; 99US-0139538P.
PR 18-JUN-1999; 99US-0139539P.
PR 18-JUN-1999; 99US-0139540P.
PR 18-JUN-1999; 99US-0139541P.
PR 18-JUN-1999; 99US-0139542P.
PR 18-JUN-1999; 99US-0139543P.
PR 18-JUN-1999; 99US-0139544P.
PR 18-JUN-1999; 99US-0139545P.
PR 18-JUN-1999; 99US-0139546P.
PR 18-JUN-1999; 99US-0139547P.
PR 18-JUN-1999; 99US-0139548P.
PR 18-JUN-1999; 99US-0139549P.
PR 18-JUN-1999; 99US-0139550P.
PR 18-JUN-1999; 99US-0139551P.
PR 18-JUN-1999; 99US-0139552P.
PR 18-JUN-1999; 99US-0139553P.
PR 18-JUN-1999; 99US-0139554P.
PR 18-JUN-1999; 99US-0139555P.
PR 18-JUN-1999; 99US-0139556P.
PR 18-JUN-1999; 99US-0139557P.
PR 18-JUN-1999; 99US-0139558P.
PR 18-JUN-1999; 99US-0139559P.
PR 18-JUN-1999; 99US-0139560P.
PR 18-JUN-1999; 99US-0139561P.
PR 18-JUN-1999; 99US-0139562P.
PR 18-JUN-1999; 99US-0139563P.
PR 18-JUN-1999; 99US-0139564P.
PR 18-JUN-1999; 99US-0139565P.
PR 18-JUN-1999; 99US-0139566P.
PR 18-JUN-1999; 99US-0139567P.
PR 18-JUN-1999; 99US-0139568P.
PR 18-JUN-1999; 99US-0139569P.
PR 18-JUN-1999; 99US-0139570P.
PR 18-JUN-1999; 99US-0139571P.
PR 18-JUN-1999; 99US-0139572P.
PR 18-JUN-1999; 99US-0139573P.
PR 18-JUN-1999; 99US-0139574P.
PR 18-JUN-1999; 99US-0139575P.
PR 18-JUN-1999; 99US-0139576P.
PR 18-JUN-1999; 99US-0139577P.
PR 18-JUN-1999; 99US-0139578P.
PR 18-JUN-1999; 99US-0139579P.
PR 18-JUN-1999; 99US-0139580P.
PR 18-JUN-1999; 99US-0139581P.
PR 18-JUN-1999; 99US-0139582P.
PR 18-JUN-1999; 99US-0139583P.
PR 18-JUN-1999; 99US-0139584P.
PR 18-JUN-1999; 99US-0139585P.
PR 18-JUN-1999; 99US-0139586P.
PR 18-JUN-1999; 99US-0139587P.
PR 18-JUN-1999; 99US-0139588P.
PR 18-JUN-1999; 99US-0139589P.
PR 18-JUN-1999; 99US-0139590P.
PR 18-JUN-1999; 99US-0139591P.
PR 18-JUN-1999; 99US-0139592P.
PR 18-JUN-1999; 99US-0139593P.
PR 18-JUN-1999; 99US-0139594P.
PR 18-JUN-1999; 99US-0139595P.
PR 18-JUN-1999; 99US-0139596P.
PR 18-JUN-1999; 99US-0139597P.
PR 18-JUN-1999; 99US-0139598P.
PR 18-JUN-1999; 99US-0139599P.
PR 18-JUN-1999; 99US-0139600P.
PR 18-JUN-1999; 99US-0139601P.
PR 18-JUN-1999; 99US-0139602P.
PR 18-JUN-1999; 99US-0139603P.
PR 18-JUN-1999; 99US-0139604P.
PR 18-JUN-1999; 99US-0139605P.
PR 18-JUN-1999; 99US-0139606P.
PR 18-JUN-1999; 99US-0139607P.
PR 18-JUN-1999; 99US-0139608P.
PR 18-JUN-1999; 99US-0139609P.
PR 18-JUN-1999; 99US-0139610P.
PR 18-JUN-1999; 99US-0139611P.
PR 18-JUN-1999; 99US-0139612P.
PR 18-JUN-1999; 99US-0139613P.
PR 18-JUN-1999; 99US-0139614P.
PR 18-JUN-1999; 99US-0139615P.
PR 18-JUN-1999; 99US-0139616P.
PR 18-JUN-1999; 99US-0139617P.
PR 18-JUN-1999; 99US-0139618P.
PR 18-JUN-1999; 99US-0139619P.
PR 18-JUN-1999; 99US-0139620P.
PR 18-JUN-1999; 99US-0139621P.
PR 18-JUN-1999; 99US-0139622P.
PR 18-JUN-1999; 99US-0139623P.
PR 18-JUN-1999; 99US-0139624P.
PR 18-JUN-1999; 99US-0139625P.
PR 18-JUN-1999; 99US-0139626P.
PR 18-JUN-1999; 99US-0139627P.
PR 18-JUN-1999; 99US-0139628P.
PR 18-JUN-1999; 99US-0139629P.
PR 18-JUN-1999; 99US-0139630P.
PR 18-JUN-1999; 99US-0139631P.
PR 18-JUN-1999; 99US-0139632P.
PR 18-JUN-1999; 99US-0139633P.
PR 18-JUN-1999; 99US-0139634P.
PR 18-JUN-1999; 99US-0139635P.
PR 18-JUN-1999; 99US-0139636P.
PR 18-JUN-1999; 99US-0139637P.
PR 18-JUN-1999; 99US-0139638P.
PR 18-JUN-1999; 99US-0139639P.
PR 18-JUN-1999; 99US-0139640P.
PR 18-JUN-1999; 99US-0139641P.
PR 18-JUN-1999; 99US-0139642P.
PR 18-JUN-1999; 99US-0139643P.
PR 18-JUN-1999; 99US-0139644P.
PR 18-JUN-1999; 99US-0139645P.
PR 18-JUN-1999; 99US-0139646P.
PR 18-JUN-1999; 99US-0139647P.
PR 18-JUN-1999; 99US-0139648P.
PR 18-JUN-1999; 99US-0139649P.
PR 18-JUN-1999; 99US-0139650P.
PR 18-JUN-1999; 99US-0139651P.
PR 18-JUN-1999; 99US-0139652P.
PR 18-JUN-1999; 99US-0139653P.
PR 18-JUN-1999; 99US-0139654P.
PR 18-JUN-1999; 99US-0139655P.
PR 18-JUN-1999; 99US-0139656P.
PR 18-JUN-1999; 99US-0139657P.
PR 18-JUN-1999; 99US-0139658P.
PR 18-JUN-1999; 99US-0139659P.
PR 18-JUN-1999; 99US-0139660P.
PR 18-JUN-1999; 99US-0139661P.
PR 18-JUN-1999; 99US-0139662P.
PR 18-JUN-1999; 99US-0139663P.
PR 18-JUN-1999; 99US-0139664P.
PR 18-JUN-1999; 99US-0139665P.
PR 18-JUN-1999; 99US-0139666P.
PR 18-JUN-1999; 99US-0139667P.
PR 18-JUN-1999; 99US-0139668P.
PR 18-JUN-1999; 99US-0139669P.
PR 18-JUN-1999; 99US-0139670P.
PR 18-JUN-1999; 99US-0139671P.
PR 18-JUN-1999; 99US-0139672P.
PR 18-JUN-1999; 99US-0139673P.
PR 18-JUN-1999; 99US-0139674P.
PR 18-JUN-1999; 99US-0139675P.
PR 18-JUN-1999; 99US-0139676P.
PR 18-JUN-1999; 99US-0139677P.
PR 18-JUN-1999; 99US-0139678P.
PR 18-JUN-1999; 99US-0139679P.
PR 18-JUN-1999; 99US-0139680P.
PR 18-JUN-1999; 99US-0139681P.
PR 18-JUN-1999; 99US-0139682P.
PR 18-JUN-1999; 99US-0139683P.
PR 18-JUN-1999; 99US-0139684P.
PR 18-JUN-1999; 99US-0139685P.
PR 18-JUN-1999; 99US-0139686P.
PR 18-JUN-1999; 99US-0139687P.
PR 18-JUN-1999; 99US-0139688P.
PR 18-JUN-1999; 99US-0139689P.
PR 18-JUN-1999; 99US-0139690P.
PR 18-JUN-1999; 99US-0139691P.
PR 18-JUN-1999; 99US-0139692P.
PR 18-JUN-1999; 99US-0139693P.
PR 18-JUN-1999; 99US-0139694P.
PR 18-JUN-1999; 99US-0139695P.
PR 18-JUN-1999; 99US-0139696P.
PR 18-JUN-1999; 99US-0139697P.
PR 18-JUN-1999; 99US-0139698P.
PR 18-JUN-1999; 99US-0139699P.
PR 18-JUN-1999; 99US-0139700P.
PR 18-JUN-1999; 99US-0139701P.
PR 18-JUN-1999; 99US-0139702P.
PR 18-JUN-1999; 99US-0139703P.
PR 18-JUN-1999; 99US-0139704P.
PR 18-JUN-1999; 99US-0139705P.
PR 18-JUN-1999; 99US-0139706P.
PR 18-JUN-1999; 99US-0139707P.
PR 18-JUN-1999; 99US-0139708P.
PR 18-JUN-1999; 99US-0139709P.
PR 18-JUN-1999; 99US-0139710P.
PR 18-JUN-1999; 99US-0139711P.
PR 18-JUN-1999; 99US-0139712P.
PR 18-JUN-1999; 99US-0139713P.
PR 18-JUN-1999; 99US-0139714P.
PR 18-JUN-1999; 99US-0139715P.
PR 18-JUN-1999; 99US-0139716P.
PR 18-JUN-1999; 99US-0139717P.
PR 18-JUN-1999; 99US-0139718P.
PR 18-JUN-1999; 99US-0139719P.
PR 18-JUN-1999; 99US-0139720P.
PR 18-JUN-1999; 99US-0139721P.
PR 18-JUN-1999; 99US-0139722P.
PR 18-JUN-1999; 99US-0139723P.
PR 18-JUN-1999; 99US-0139724P.
PR 18-JUN-1999; 99US-0139725P.
PR 18-JUN-1999; 99US-0139726P.
PR 18-JUN-1999; 99US-0139727P.
PR 18-JUN-1999; 99US-0139728P.
PR 18-JUN-1999; 99US-0139729P.
PR 18-JUN-1999; 99US-0139730P.
PR 18-JUN-1999; 99US-0139731P.
PR 18-JUN-1999; 99US-0139732P.
PR 18-JUN-1999; 99US-0139733P.
PR 18-JUN-1999; 99US-0139734P.
PR 18-JUN-1999; 99US-0139735P.
PR 18-JUN-1999; 99US-0139736P.
PR 18-JUN-1999; 99US-0139737P.
PR 18-JUN-1999; 99US-0139738P.
PR 18-JUN-1999; 99US-0139739P.
PR 18-JUN-1999; 99US-0139740P.
PR 18-JUN-1999; 99US-0139741P.
PR 18-JUN-1999; 99US-0139742P.
PR 18-JUN-1999; 99US-0139743P.
PR 18-JUN-1999; 99US-0139744P.
PR 18-JUN-1999; 99US-0139745P.
PR 18-JUN-1999; 99US-0139746P.
PR 18-JUN-1999; 99US-0139747P.
PR 18-JUN-1999; 99US-0139748P.
PR 18-JUN-1999; 99US-0139749P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139751P.
PR 18-JUN-1999; 99US-0139752P.
PR 18-JUN-1999; 99US-0139753P.
PR 18-JUN-1999; 99US-0139754P.
PR 18-JUN-1999; 99US-0139755P.
PR 18-JUN-1999; 99US-0139756P.
PR 18-JUN-1999; 99US-0139757P.
PR 18-JUN-1999; 99US-0139758P.
PR 18-JUN-1999; 99US-0139759P.
PR 18-JUN-1999; 99US-0139760P.
PR 18-JUN-1999; 99US-0139761P.
PR 18-JUN-1999; 99US-0139762P.
PR 18-JUN-1999; 99US-0139763P.
PR 18-JUN-1999; 99US-0139764P.
PR 18-JUN-1999; 99US-0139765P.
PR 18-JUN-1999; 99US-0139766P.
PR 18-JUN-1999; 99US-0139767P.
PR 18-JUN-1999; 99US-0139768P.
PR 18-JUN-1999; 99US-0139769P.
PR 18-JUN-1999; 99US-0139770P.
PR 18-JUN-1999; 99US-0139771P.
PR 18-JUN-1999; 99US-0139772P.
PR 18-JUN-1999; 99US-0139773P.
PR 18-JUN-1999; 99US-0139774P.
PR 18-JUN-1999; 99US-0139775P.
PR 18-JUN-1999; 99US-0139776P.
PR 18-JUN-1999; 99US-0139777P.
PR 18-JUN-1999; 99US-0139778P.
PR 18-JUN-1999; 99US-0139779P.
PR 18-JUN-1999; 99US-0139780P.
PR 18-JUN-1999; 99US-0139781P.
PR 18-JUN-1999; 99US-0139782P.
PR 18-JUN-1999; 99US-0139783P.
PR 18-JUN-1999; 99US-0139784P.
PR 18-JUN-1999; 99US-0139785P.
PR 18-JUN-1999; 99US-0139786P.
PR 18-JUN-1999; 99US-0139787P.
PR 18-JUN-1999; 99US-0139788P.
PR 18-JUN-1999; 99US-0139789P.
PR 18-JUN-1999; 99US-0139790P.
PR 18-JUN-1999; 99US-0139791P.
PR 18-JUN-1999; 99US-0139792P.
PR 18-JUN-1999; 99US-0139793P.
PR 18-JUN-1999; 99US-0139794P.
PR 18-JUN-1999; 99US-0139795P.
PR 18-JUN-1999; 99US-0139796P.
PR 18-JUN-1999; 99US-0139797P.
PR 18-JUN-1999; 99US-0139798P.
PR 18-JUN-1999; 99US-0139799P.
PR 18-JUN-1999; 99US-0139800P.
PR 18-JUN-1999; 99US-0139801P.
PR 18-JUN-1999; 99US-0139802P.
PR 18-JUN-1999; 99US-0139803P.
PR 18-JUN-1999; 99US-0139804P.
PR 18-JUN-1999; 99US-0139805P.
PR 18-JUN-1999; 99US-0139806P.
PR 18-JUN-1999; 99US-0139807P.
PR 18-JUN-1999; 99US-0139808P.
PR 18-JUN-1999; 99US-0139809P.
PR 18-JUN-1999

Qy 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 1064 ATAAACCGGGCAGTCTCATCGATGTTGGGATTAATCTTTCAGGATCCAGTGCT 1123
Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrCysGluAlaSerLysIleAlaGly 260
Db 1124 GCGGTGGATGATGATGTTGGAGCAATTTGCTACGAGGAGCTAGCAAGTTGCATCA 1183
Qy 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 1184 GCCATCACCTGTCTCTGCGGTGAGGACCAATGACCATGACCATGCTTCTATCCAAC 1243
Qy 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1244 ACTTTAATCATCAGTAAGAGGATTCACAACCTCCAG 1279
RESULT 5
ID AAC34458
XX AAC34458 standard; DNA; 1066 BP.
AC AAC34458;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 6720.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
PD
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127452P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135829P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 29-JUN-1999; 99US-0140823P.
PR 30-JUN-1999; 99US-0140991P.
PR 01-JUL-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.

23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154799P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156598P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159631P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 1,158-104
Score: 1035.00
Percent Similarity: 82.99%
Best Local Similarity: 67.36%
Query Match: 69.65%
DB: 3
Length: 1182
Matches: 194
Conservative: 45
Mismatch: 49
Indels: 0
Gaps: 0

US-10-723-061-18 (1-292) x AAC39538 (1-1182)

QY 1 AlaLysIleAspGlyLysLeuAlaLysGlnIleArgGluGluAlaValGlu 20
DB 114 GCGAAGATATCCAGCGGAGGCGATTGCTATACATCAGATCAGATCCGAGGAA 173
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
DB 174 GTTCGGGTCTATCTGAGAAACACGCAAGGTCGCCAGGACTAGCTGTATTATTGTTGGG 233

41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
234 AGCCGAAGGATTTCACAGACCTATGTGAATACGAGAGGAGGCGTGCCTGAGGTTGGG 293
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
294 ATTAAGTCATTGACGTGGCCTACAGAGGAGGTTCTGAAGCTGATCTTATTAGCAA 353
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
354 GTTCATGAACTAATCAATCCGATGTCCTGATGATAGTCACTCCATTCCTGCGG 413
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
414 AAACATATTATGAGGAGATATATTGGGTGCAATCAGCATTTGATTAAGATGTCGATGGC 473
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
474 TTCCATCCCTTGAATATTGTTAGCTAGCCATGAAGAGGAGAGAACCCCTCTCTCTCT 533
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
534 TGCACCCCAAGGGATGTTTGAACCTCTAGCTAGAGGCGGCTAAAGATAAAGGGGCAA 593
QY 161 ArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
594 CGAGCAGTTGTTAGTTCGAGTAACATTTGTTGTTGCTTCCCTTCTCTCTCTCTCTCT 653
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
654 AAGGCTGATGCTACTGTCAACACTGTACATCTCTCACCAAGGATCCTGAGGCTATCATA 713
QY 201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
714 CGGGAAGCTGACATTTGTTGCTGATCGGCAAGCCACATGATTAAGGCAACTGG 773
QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
774 ATAAAGCCAGGCTGACGATATTGATTTGGAACCTAATGCAGTCAGCGACCCGAGCAAG 833
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
834 AAATCAGGATACCGTTGGTTGGAGATGTTGATTTCCGAGAAGCTTCAAAAGTTGAGGT 893
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
894 TTCATTAACCTCCGTCCTGTTGTTGAGGCCCATGACAGTGGCAATGCTTCTCAGGAAC 953
QY 281 ThrLeuGluSerAlaLysArgile 288
954 ACCGTAGACGTCGCAAGGCTGTC 977

RESULT 7

AAC45046

ID AAC45046 standard; DNA; 1688 BP.

XX AAC45046;

XX AC

XX 18-OCT-2000 (first entry)

XX DT

XX XX

XX DE

XX XX

XX Hybridisation assay; genetic mapping; gene expression control;

XX KW

XX protein identification; signal transduction pathway; metabolic pathway;

XX KW

XX promoter; termination sequence; ss.

XX OS

XX Arabidopsis thaliana.

XX PN

XX EP1033405-A2.

XX XX

XX PD

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121025P. PR 12-JUL-1999; 99US-0142977P.
PR 05-MAR-1999; 99US-0123180P. PR 13-JUL-1999; 99US-0143542P.
PR 09-MAR-1999; 99US-0123548P. PR 14-JUL-1999; 99US-0143624P.
PR 23-MAR-1999; 99US-0125788P. PR 15-JUL-1999; 99US-0144005P.
PR 25-MAR-1999; 99US-0126264P. PR 16-JUL-1999; 99US-0144085P.
PR 28-MAR-1999; 99US-0126785P. PR 16-JUL-1999; 99US-0144086P.
PR 01-APR-1999; 99US-0127462P. PR 19-JUL-1999; 99US-0144325P.
PR 08-APR-1999; 99US-0128234P. PR 19-JUL-1999; 99US-0144331P.
PR 08-APR-1999; 99US-0128714P. PR 19-JUL-1999; 99US-0144332P.
PR 16-APR-1999; 99US-0129845P. PR 19-JUL-1999; 99US-0144333P.
PR 19-APR-1999; 99US-0130077P. PR 19-JUL-1999; 99US-0144334P.
PR 21-APR-1999; 99US-0130449P. PR 20-JUL-1999; 99US-0144335P.
PR 23-APR-1999; 99US-0130510P. PR 20-JUL-1999; 99US-0144352P.
PR 28-APR-1999; 99US-0130891P. PR 20-JUL-1999; 99US-0144632P.
PR 30-APR-1999; 99US-0131449P. PR 20-JUL-1999; 99US-0144884P.
PR 30-APR-1999; 99US-0132048P. PR 21-JUL-1999; 99US-0144814P.
PR 04-MAY-1999; 99US-0132407P. PR 21-JUL-1999; 99US-0145086P.
PR 05-MAY-1999; 99US-0132484P. PR 21-JUL-1999; 99US-0145088P.
PR 06-MAY-1999; 99US-0132485P. PR 22-JUL-1999; 99US-0145085P.
PR 06-MAY-1999; 99US-0132487P. PR 22-JUL-1999; 99US-0145087P.
PR 07-MAY-1999; 99US-0132863P. PR 22-JUL-1999; 99US-0145089P.
PR 11-MAY-1999; 99US-0134256P. PR 22-JUL-1999; 99US-0145192P.
PR 14-MAY-1999; 99US-0134218P. PR 23-JUL-1999; 99US-0145145P.
PR 14-MAY-1999; 99US-0134219P. PR 23-JUL-1999; 99US-0145218P.
PR 14-MAY-1999; 99US-0134321P. PR 26-JUL-1999; 99US-0145224P.
PR 18-MAY-1999; 99US-0134370P. PR 27-JUL-1999; 99US-0145913P.
PR 18-MAY-1999; 99US-0134768P. PR 27-JUL-1999; 99US-0145918P.
PR 19-MAY-1999; 99US-0134941P. PR 27-JUL-1999; 99US-0145919P.
PR 20-MAY-1999; 99US-0135124P. PR 28-JUL-1999; 99US-0145951P.
PR 21-MAY-1999; 99US-0135353P. PR 02-AUG-1999; 99US-0146386P.
PR 24-MAY-1999; 99US-0135629P. PR 02-AUG-1999; 99US-0146388P.
PR 25-MAY-1999; 99US-0136021P. PR 02-AUG-1999; 99US-0146389P.
PR 28-MAY-1999; 99US-0136782P. PR 03-AUG-1999; 99US-0147038P.
PR 01-JUN-1999; 99US-0137222P. PR 04-AUG-1999; 99US-0147204P.
PR 03-JUN-1999; 99US-0137528P. PR 04-AUG-1999; 99US-0147302P.
PR 04-JUN-1999; 99US-0137502P. PR 05-AUG-1999; 99US-0147192P.
PR 07-JUN-1999; 99US-0137724P. PR 05-AUG-1999; 99US-0147260P.
PR 08-JUN-1999; 99US-0138094P. PR 06-AUG-1999; 99US-0147303P.
PR 10-JUN-1999; 99US-0138540P. PR 06-AUG-1999; 99US-0147416P.
PR 10-JUN-1999; 99US-0138847P. PR 09-AUG-1999; 99US-0147493P.
PR 14-JUN-1999; 99US-0139119P. PR 09-AUG-1999; 99US-0147935P.
PR 16-JUN-1999; 99US-0139452P. PR 10-AUG-1999; 99US-0148171P.
PR 16-JUN-1999; 99US-0139453P. PR 11-AUG-1999; 99US-0148319P.
PR 17-JUN-1999; 99US-0139492P. PR 12-AUG-1999; 99US-0148341P.
PR 18-JUN-1999; 99US-0139454P. PR 13-AUG-1999; 99US-0148565P.
PR 18-JUN-1999; 99US-0139455P. PR 13-AUG-1999; 99US-0148684P.
PR 18-JUN-1999; 99US-0139456P. PR 16-AUG-1999; 99US-0149368P.
PR 18-JUN-1999; 99US-0139457P. PR 17-AUG-1999; 99US-0149175P.
PR 18-JUN-1999; 99US-0139458P. PR 18-AUG-1999; 99US-0149426P.
PR 18-JUN-1999; 99US-0139459P. PR 20-AUG-1999; 99US-0149722P.
PR 18-JUN-1999; 99US-0139460P. PR 20-AUG-1999; 99US-0149723P.
PR 18-JUN-1999; 99US-0139461P. PR 20-AUG-1999; 99US-0149929P.
PR 18-JUN-1999; 99US-0139462P. PR 23-AUG-1999; 99US-0149902P.
PR 18-JUN-1999; 99US-0139463P. PR 23-AUG-1999; 99US-0149930P.
PR 18-JUN-1999; 99US-0139464P. PR 25-AUG-1999; 99US-0150566P.
PR 18-JUN-1999; 99US-0139465P. PR 26-AUG-1999; 99US-0150884P.
PR 18-JUN-1999; 99US-0139466P. PR 26-AUG-1999; 99US-0150884P.
PR 18-JUN-1999; 99US-0139467P. PR 27-AUG-1999; 99US-0151065P.
PR 18-JUN-1999; 99US-0139468P. PR 27-AUG-1999; 99US-0151066P.
PR 18-JUN-1999; 99US-0139469P. PR 27-AUG-1999; 99US-0151080P.
PR 18-JUN-1999; 99US-0139470P. PR 27-AUG-1999; 99US-0151080P.
PR 18-JUN-1999; 99US-0139471P. PR 30-AUG-1999; 99US-0151303P.
PR 18-JUN-1999; 99US-0139472P. PR 31-AUG-1999; 99US-0151338P.
PR 18-JUN-1999; 99US-0140353P. PR 01-SEP-1999; 99US-0151930P.
PR 18-JUN-1999; 99US-0140354P. PR 07-SEP-1999; 99US-0152363P.
PR 18-JUN-1999; 99US-0140695P. PR 13-SEP-1999; 99US-0153070P.
PR 18-JUN-1999; 99US-0140823P. PR 15-SEP-1999; 99US-0153758P.
PR 18-JUN-1999; 99US-0140951P. PR 16-SEP-1999; 99US-0154018P.
PR 18-JUN-1999; 99US-0141287P. PR 16-SEP-1999; 99US-0154039P.
PR 18-JUN-1999; 99US-0141842P. PR 20-SEP-1999; 99US-0154779P.
PR 18-JUN-1999; 99US-0142154P. PR 22-SEP-1999; 99US-0155139P.
PR 18-JUN-1999; 99US-0142055P. PR 23-SEP-1999; 99US-0155486P.
PR 18-JUN-1999; 99US-0142390P. PR 24-SEP-1999; 99US-0155659P.
PR 18-JUN-1999; 99US-0142803P. PR 28-SEP-1999; 99US-0156458P.
PR 18-JUN-1999; 99US-0142920P.

```

PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158389P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162143P.

Alignment Scores:
Pred. No.: 1,866-102 Length: 1688
Score: 1017.00 Matches: 193
Percent Similarity: 80.07% Conservative: 36
Best Local Similarity: 67.48% Mismatches: 57
Query Match: 68.44% Indels: 0
DB: 3 Gaps: 0

US-10-723-061-18 (1-292) x AAC45046 (1-1688)

QY 3 llelleAspGlyLysLeuValAlaLysGlnlleArgGluGluLeuAlaValGluLeuAla 22
DB 631 GTGATAGATGGTAATGTTATCTCCTGAGGAATCAGAACAAAGATTATTAGTGAAGTCGGG 690

QY 23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValleLeuValGlySerArg 42
DB 691 AAGATGAAAAGGCTGTGGAAAAGTCCCTGCTGCTGTTGTTGTTGTTGGTGAACAA 750

QY 43 LysAspSerGlnThrTyValArgAsnLysLysLysAlaCysGluAlaValGlyLeuLys 62
DB 751 AGAGACTGCCAAACTATGTTGCCAACAAATTAAGCTTGTGAAGAACTGGCATTA 810

QY 63 SerTyGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
DB 811 TCTGTTTGGCTGAATACACAGGATGTACTGAAGCAGACAGATTATCAGTGTATTGAGA 870

QY 83 ThrPheAsnSerAspProSerValHisGlylleLeuValGlnLeuProLeuProHisHis 102
DB 871 AAATTCAATGAAGACACGCTATTACCGAATTCCTGTGCAACTTCCTTTACCCCAACAT 930

QY 103 MetAspGluAsnLysLeuAsnAlaValSerlleGluLysAspValAspGlyPheHis 122
DB 931 CTTATGATCAAAATCTTGATATGTTGATGATGGAGAAAGATGTTGATGGGTTTCAT 990

QY 123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
DB 991 CCATTAAACGTGGGAATCTTCAATGATGAGGGGGAACCGCTGTTGTATCTTGTCACT 1050

```

```

QY 143 ProLysGlyCysMetGluLeuLeuHisArgTyGlyValGluLeuLysGlyLysArgAla 162
DB 1051 CCTAAGGGCTGTGTGGAGTTTGTGATAGAACGGGTGTTGAAATAGCAGGAAAAATGCT 1110

QY 163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLysAla 182
DB 1111 GTGGTCATTAGGAGAGACACACATCGTTGGACTACCAATGCTCTTTATTATTCAGAGGCAT 1170

QY 183 AsnAlaThrValSerlleValHisSerAsnThrLysLysProGluGluLeuThrArgGln 202
DB 1171 GACCGCACAGTAAGCACCGGTGTCATCGCTTACAAAGGATCCAGAGCACATTACTAGGAAG 1230

QY 203 AlaAspIleValleAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
DB 1231 GCGGATATCCTTATACACACGCGGTATACCTAATTAGTTCGGAAGTTGGCTTAAG 1290

QY 223 ProGlyAlaAlaIleleAspValGlylleAsnProValAspAspProGluSerProArg 242
DB 1291 CCTGGAGCAGTTGTCTATTGATGTTGGACAACTCCAGTCGAGGATAGTAGCTGTGAGTTT 1350

QY 243 GlyTyArgLeuValGlyAspValCysTyGluGluAlaSerLysIleleAlaGlyLeuIle 262
DB 1351 GGTACCGTCTTGTCCGGGATGTATGTACGAGGAGCGTTAGGTGTTCCCTCGGCTATC 1410

QY 263 ThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282
DB 1411 ACTCCGTACCGGAGGAGTGTGACCTATGACGATCACCATGCTACTATGCAATCTTGG 1470

QY 283 GluSerAlaLysArgIle 288
DB 1471 GAGGCTGCTAAGCGGATC 1488

RESULT 8
AAC36669
ID AAC36669 standard; DNA; 1327 BP.
XX
AC AAC36669;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 14649.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
PN EF1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.

```



```
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 4,766-102 Length: 1327
Score: 1012.00 Matches: 192
Percent Similarity: 80.07% Conservaive: 37
Best Local Similarity: 67.13% Mismatches: 57
Query Match: 68.10% Indels: 0
DB: 3 Gaps: 0

US-10-723-061-18 (1-292) x AAC36669 (1-1327)

QY 3 IlelleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
DB 270 GTGATAGATGGTAATGTTATCGCTGAGGAATCAGAACAAAGATTATTAGTGAAGTGGG 329

QY 23 LysMetLysAspAlaIleGlyValProGlyLeuAlaValIleLeuValGlySerArg 42
DB 330 AAGATGAAAGAGCGTGTGGAAAGTCCCTGCTGCTGTTGTTGGTTGGTGAACAA 389

QY 43 LysAspSerGlnThrTyrrValargAsnLysLysAlaCysGluAlaValGlyIleLys 62
DB 390 AGAGACTCGCAACTATGTTGCGAACAAATAAAGCTTGTGAAGAACTGGCATTA 449

QY 63 SerTyrrGluValAsnLeuProGluAspSerGluAspGluValLeuLysHisIleAla 82
DB 450 TCTGTTTGGCTGAATACAGAGATGTACTGAGGACAGATTATCAGTGTATTGAGA 509

QY 83 ThrPheAsnSerAspProSerValHisGlyLeuValGlnLeuProLeuProHisHis 102
DB 510 AAATTCATGAAGACACAGCTATTACCGAATCTCTGTGCAACTTCCTTTACCCCAACAT 569

QY 103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
DB 570 CTGTGATGAATCAAAATCTTGATATGGTGAGATTGGAGAAAGATGTTGATGGGTTTCAT 629

QY 123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
DB 630 CCATTAAACGTGGGGAATCTTCAATGAGAGGGAGGAAACCGCTGTTGTGCTTGCAC 689

QY 143 ProLysGlyCysMetGluLeuLeuHisArgTyrrGlyValGluIleLysGlyLysArgAla 162
DB 690 CCTAAGGGCTGTGTGGAGTGTGTGTAAGAACGGGTGTTGAAATAGCAGGAAAAAATGCT 749

QY 163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLysAla 182
DB 750 GTGGTCATAGGCAGAGCAACATCGTTGCTACCAATGCTCTTTATTATTCAGAGGCAT 809

QY 183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
DB 810 GACGCGACAGTAAGCAGCGGTGCTGCTGTTTACAAAGGATCCAGAGCACATTACTAGGAAG 869

QY 203 AlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
DB 870 GCGGATATCGTTATACAGACGCGGTATACCTAATTTAGTTCGTGGAAGTTGGTGAAG 929

QY 223 ProGlyValAlaIleAspValGlyIleAsnProValAspProGluSerProArg 242
```


Db 90813 GAAATGCTACAGTAAACAATAGCTCATACCGGTACAAAAGATTACCTCAAGTGGCAAAA 90872
Qy 202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
Db 90873 GAAGCCGATATTCTAGTTGTTGCAACAGGATTAGCTAAATTTGTTAAAAAGAAATACATT 90932
Qy 222 LysProGlyAlaAlaIleIleAspValGlyIleAspValAspProGluSerPro 241
Db 90933 AAGCAGCGCAATGTTATTGATGTTGGTATG-----GACCAGCGAT 90974
Qy 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
Db 90975 GAAAAACAATAATTATGTGGCGATGTAGATTTCGATGATGTCAAAGAGCAGCAGGATTC 91034
Qy 262 IleThrProValProGlyGlyValCysProMetThrIleAlaMetLeuLeuSerAsnThr 281
Db 91035 ATCACTCCAGTTCAGAGGCGTTGGGCCAATGACAATTACCATGCTCTTGTCTTAATACA 91094
Qy 282 LeuGluSerAlaLysArgIleHisLys 290
Db 91095 TTAAAGCTGCAAAACGCATTGGAG 91121
RESULT 11
ABQ67195_3/c
Continuation (4 of 5) of ABQ67195 from base 300001 (Listeria innocua contig DNA sequence
WP Sequence split into 5 fragments LOCUS ABQ67195 Accession Abq67195
WP Fragment Name Begin End
WP ABQ67195_0 1 110000
WP ABQ67195_1 100001 210000
WP ABQ67195_2 200001 310000
WP ABQ67195_3 300001 410000
WP ABQ67195_4 400001 495269
Alignment Scores:
Pred. No.: 7,84e-72 Length: 110000
Score: 763.50 Matches: 151
Percent Similarity: 72.6% Conservative: 59
Best Local Similarity: 52.2% Mismatches: 70
Query Match: 51.3% Indels: 9
DB: 6 Gaps: 4
US-10-723-061-18 (1-292) x ABQ67195_3 (1-110000)
Qy 2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
Db 68669 GAAATATTGATGGCAAAAAGTTATCCCAAGAAATCCCAAGAGAAAGTAACTACATCAAGATGA 68610
Qy 22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
Db 68609 GCTGAATTAGTAAACAA---GGTAAAAAACCGGCTCTTGGCTTGTCTCGTAGGAGAT 68553
Qy 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
Db 68552 AACCAAGCCTCTCGTACATATCTTAGAATAAACAAAAGCCGAACAGAAAGACGAGGAATG 68493
Qy 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
Db 68492 AAATCTGTTTAAATGAATCTTCTGAAACAGTAACAGAGAAAGAAATTAATAAGTAGTA 68433
Qy 82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
Db 68432 GAAGATTAATCAAGATAATAACAATTCATGATATTGGTACATTAACCTTTACCAAAA 68373
Qy 102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
Db 68372 CATATTTCAGAGAAAAGATCAATTGATGCTATTAGTTTGCACAAAGATGTTGACCGTTTC 68313
Qy 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
Db 68312 CACCCGGTGAATAGTAGTAACATTATTCAT---GGTAAGATCT---TTGCTTCGTCG 68259
Qy 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161

Db 68258 ACACCGCAGGAATATTAGTTAATTAATCAACGGCACTCAATAGAGCAACCA 68199
Qy 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLys 181
Db 68198 GCAGTCGTTATTGGCAGAGTAATATTGAGGTAAAGCCAGTAGCACAATTATTATAAT 68139
Qy 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
Db 68138 GAAATGCTACAGTAAACATAGCTCATAGCCGTACAAAAGATTACCTCAAGTGGCAAAA 68079
Qy 202 GlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
Db 68078 GAAGCCGATATTCTAGTTGTTGTCACAGGATAGCTAAATTTGTTAAAAAGAAATACATT 68019
Qy 222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspProGluSerPro 241
Db 68018 AAGCCAGGCGCAATGTTATTGATGTTGGTATG-----GACCAGCGAT 67977
Qy 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
Db 67976 GAAAAACAATAATTATGTGGCGATGTAGATTTTGATGATGTCAAGAGCAAGCAGGATTC 67917
Qy 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
Db 67916 ATCACTCCAGTTCAGAGGCGTTGGGCCAATGACAATTACCATGCTTCTTGTAAATACA 67857
Qy 282 LeuGluSerAlaLysArgIleHisLys 290
Db 67856 TTAAAGCTGCAAAACGCATTGGAG 67830
RESULT 12
ABL07155
ID ABL07155 standard; cDNA; 3003 BP.
XX ABL07155;
XX AC ABL07155;
XX XX
DT 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 15947.
XX XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX XX
KW pharmaceutical; gene; ss.
XX XX
OS Drosophila melanogaster.
XX XX
PN WO200171042-A2.
XX XX
PD 27-SEP-2001.
XX XX
PF 23-MAR-2001; 2001WO-US009231.
XX XX
PR 23-MAR-2000; 2000US-0191637P.
XX XX
PR 11-JUL-2000; 2000US-00614150.
XX XX
PA (PEKE) PE CORP NY.
XX XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX XX
DR WPI; 2001-656860/75.
XX XX
DR P-PSDB; ABB63052.
XX XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX XX
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX XX
XX interactions.
XX XX
PS Claim 1; SEQ ID NO 15947; 21pp + Sequence Listing; English.
XX XX
CC The invention relates to an isolated nucleic acid detection reagent
XX XX
CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX XX
CC useful in developmental biology and in elucidating cell signalling and
XX XX
CC cell-cell interactions in higher eukaryotes for the development of
XX XX
CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX XX
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB572072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3003 BP; 598 A; 798 C; 889 G; 618 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 7.59e-74 Length: 3003
 Score: 762.00 Matches: 160
 Percent Similarity: 68.28% Conservative: 38
 Best Local Similarity: 55.17% Mismatches: 88
 Query Match: 51.28% Indels: 4
 DB: 4 Gaps: 3

US-10-723-061-18 (1-292) x ABL07155 (1-3003)

QY 1 AlalysileleAspGlyLysLeuValAlaLysGlnIleArgGluLileAlaValGlu 20
 Db 208 GCCAAGATCATTCGGGCACGGCGTGGCCAAATCCATTCGGAAGAGCTGCGTAACGAG 267
 QY 21 IleAlalysMet---LysAspAlaIleGlyValProGlyLeuAlaValIleVal 39
 Db 268 GTGACGGCTATGACCAAGCAATTGGCGGATTGTCGCCGGCTTAAGGATTGTCAGGTT 327
 QY 40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaVal 59
 Db 328 GTTGAGCGTGGAGACTCCAAAGTTTACATCCGTCATGAAGATCAAGCGCGCTACGGAGATC 387
 QY 60 GlyLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
 Db 388 GGTATCGATGCCGCCACGTCACAGTCCAGTCCCGCATCCATCCAGGAGTGGATGCTCGAT 447
 QY 80 HisIleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeu 99
 Db 448 AAGATAAACGATCTGAACGAGGACCCCGGGTGCACGGCATCATAGTGCATAATGCCCTG 507
 QY 100 -----ProHisMetAsnAspGluAsnIleLeuAlaValSerIleGluLysAsp 117
 Db 508 GATTGTGACACACCCATCGATTCCGACCGAATTACGGACGCGCTTCCCGGAGAGGAT 567
 QY 118 ValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhe 137
 Db 568 GTTGACGGCTTGACACAGGCTGTGTTGAGAGTTCGCTGGCCATC---GGTGACCTGGTGT 624
 QY 138 PheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIle 157
 Db 625 TTCTGCGCTGCACTCTTGGGGTTGCTGGAGTTAATTCGGCTTCTGGAGTAGAGATC 684
 QY 158 LysGlyLysArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeu 177
 Db 685 GCCGAGACCGAGGCTGTGTTGAGAGTTCGAGACGACAGATCGTGGCAGCTCCCGCCGCGAA 744
 QY 178 LeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGlu 197
 Db 745 CTGCTCAAGTGGCCCAATGCCAGGTCACAGTTCGCTCAAAAGACCGCGACCTGGAG 804
 QY 198 GluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArg 217
 Db 805 GAGATCACCGAAGTSCAGACATCTGGTGGTGGAGTTGGAGTGGCGAGATGTTAAG 864
 QY 218 GlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAsp 237
 Db 865 GGATCGTGATATAAGCCCGAGCAGTGGTCATCGATTGCGGCATCAATGTGAAGCCAGAC 924
 QY 238 ProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLys 257
 Db 925 GCCAGCAGCCAGTGGCAGCAGCTGTGTCGGGAGCTGGACTACCGGAGCTCTCCAA 984
 QY 258 IleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeu 277
 Db 985 GTGGCAGGACATTGACGCCAGTGGCCGCGCGTGGGACCATGACCGTGGCCATGCTT 1044

QY 278 LeuSerAsnThrLeuGluSerAlaLysArg 287
 Db 1045 ATGAAAAACACCGTGGGATCTGGGCCCGA 1074

RESULT 13

ABA03041_13

Continuation (14 of 30) of ABA03041 from base 1300001 (Listeria monocytogenes EGD-e genom
 WP Sequence split into 30 fragments LOCUS ABA03041 Accession ABA03041

WP Fragment Name Begin End
 WP ABA03041_00 1 110000
 WP ABA03041_01 100001 210000
 WP ABA03041_02 200001 310000
 WP ABA03041_03 300001 410000
 WP ABA03041_04 400001 510000
 WP ABA03041_05 500001 610000
 WP ABA03041_06 600001 710000
 WP ABA03041_07 700001 810000
 WP ABA03041_08 800001 910000
 WP ABA03041_09 900001 1010000
 WP ABA03041_10 1000001 1110000
 WP ABA03041_11 1100001 1210000
 WP ABA03041_12 1200001 1310000
 WP ABA03041_13 1300001 1410000
 WP ABA03041_14 1400001 1510000
 WP ABA03041_15 1500001 1610000
 WP ABA03041_16 1600001 1710000
 WP ABA03041_17 1700001 1810000
 WP ABA03041_18 1800001 1910000
 WP ABA03041_19 1900001 2010000
 WP ABA03041_20 2000001 2110000
 WP ABA03041_21 2100001 2210000
 WP ABA03041_22 2200001 2310000
 WP ABA03041_23 2300001 2410000
 WP ABA03041_24 2400001 2510000
 WP ABA03041_25 2500001 2610000
 WP ABA03041_26 2600001 2710000
 WP ABA03041_27 2700001 2810000
 WP ABA03041_28 2800001 2910000
 WP ABA03041_29 2900001 2944528

Alignment Scores:

Pred. No.: 6.04e-71 Length: 110000
 Score: 755.50 Matches: 149
 Percent Similarity: 72.32% Conservative: 60
 Best Local Similarity: 51.56% Mismatches: 71
 Query Match: 50.84% Indels: 9
 DB: 6 Gaps: 4

US-10-723-061-18 (1-292) x ABA03041_13 (1-110000)

QY 2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluLileAlaValGluIle 21
 Db 83495 GAAATTATTGTCGCAAAAAGTTAGCAAAAGAAATTCAGAAAAGTAAACAGAGAGTA 83554
 QY 22 AlalysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
 Db 83555 GCTGAATTAGTAAAGAAA---GGTAAACACCTGCTGCTTCTGCTCGTAGCGAC 83611
 QY 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGlyIle 61
 Db 83612 AATCAACATCTCGTACATATGTAAAGAAATTAACAAAAGAACAGAGACCGGTATG 83671
 QY 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
 Db 83672 AAATCTGTTTAAATGAACTCCAGAAAATGTAAAGAAAATTAATCTATCTGTCGTA 83731
 QY 82 AlatrPheAsnSerProSerValHisGlyIleValGlnLeuProLeuProHis 101
 Db 83732 GAGAGCTAAATGAAGATAAAACCATTCACGGATACCTGTTGAGTTACCTACCAGAA 83791
 QY 102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
 Db 83792 CACATTTACAGAGAAAAGTAATTGATCTATTAGCTTTGACAAAGATGTTGACGGTTTT 83851

Qy 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCys 141
Db 83852 CATCCAGTGAATGATGTAATTTATTCATT---GGAAAGATTCA---TTTGTCCTTGT 83905
Qy 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
Db 83906 ACACGAGTGAATATTGAACTTTATTAATAAACAACCGCACTCAATAGAGGTAACGC 83965
Qy 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaLeuLeuGlnLys 181
Db 83966 GCTGTCGTTATTGTAGAGTAATATCTAGGAAACACGAGTACCCCAATCTCTTAAT 84025
Qy 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
Db 84026 GAAACCGGACAGTAACCATTCGCGATAGCCGTACAAAGATTATACCAAGTAGCGAAA 84085
Qy 202 GlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
Db 84086 GAAGCGGATATCTTGTTGTAGCACAGTTTAGCTAAATTTGTGAAAGAACTATATC 84145
Qy 222 LysProGlyAlaAlaIleAlaIleAspValGlyIleAsnProValAspAspProGluSerPro 241
Db 84146 AAACCGAGTGGATTGTTATTGATGTTGCGATG-----GATCGCGAC 84187
Qy 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
Db 84188 GAAATAATAAGTTATGTCGGTGATGTTGACCTTGATGACGTAGTACAGAGCGGATTC 84247
Qy 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
Db 84248 ATTACGCCAGTACCAGTGGCGTGGCCCGATGACATCACATGCTACTTGCGAACACA 84307
Qy 282 LeuGluSerAlaLysArgIleHisLys 290
Db 84308 TTAAGCAGCAAAACGCAATTTGAAA 84334

RESULT 14
ABQ71010
ID ABQ71010 standard; DNA; 5635 BP.
XX
AC ABQ71010;
DT 29-AUG-2003 (revised)
DT 29-AUG-2002 (first entry)

XX Listeria monocytogenes 4b contig DNA sequence #952.
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX Listeria monocytogenes ATCC 19115.
XX WO200228891-A2.
XX 11-APR-2002.
XX 04-OCT-2001; 2001WO-FR003061.
XX 04-OCT-2000; 2000FR-00012697.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators.
XX Claim 14; SEQ ID NO 3823; 180pp; French.

XX The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC on 29-AUG-2003 to standardise OS field)

XX
SQ Sequence 5635 BP; 1896 A; 927 C; 1238 G; 1574 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2 66e-72 Length: 5635
Score: 751.50 Matches: 148
Percent Similarity: 73.01% Conservative: 63
Best Local Similarity: 51.21% Mismatches: 69
Query Match: 50.57% Indels: 9
DB: 6 Gaps: 4

US-10-723-061-18 (1-292) x ABQ71010 (1-5635)

Qy 2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
Db 439 GAAATATTATTGATGCAAAAGATTAGCAAAAGAAATTTCAAGAAAAAGTAAACAAGAGA 498
Qy 22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
Db 499 GCTGAATTAGTAAAGAA---GGTAGAAACACGAGGCTTGTCTGTGCTGCTGGCGAC 555
Qy 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
Db 556 AATCAAGCATCTCGTACATATGTAGAAATAAACAACAAACGAGACAGAGCGGGATG 615
Qy 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
Db 616 AAATCCGTTTAAATTGAACCTTCAGAAATGTAAACAGAAAGAAATTTACTATCTGTGTA 675
Qy 82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
Db 676 GAGGAGCTTAACGAAGATAAACTATTTCATGTCATCTCGTCAGTTACCACTACCAGAA 735
Qy 102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
Db 736 CACATTTCAAGAAAAAGTAATTTGATCTATTAGTATGACAAAGATGTTGACGGTTTC 795
Qy 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCys 141
Db 796 CATCCAGTGAATGATGTAATTTATTCATC---GGAAAGATTCT---TTTGTCCTTGT 849
Qy 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
Db 850 ACACGAGCAAGTAATTTGAACCTTATAAATACCGCGCACTCAATAGAGCAACGC 909
Qy 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaLeuLeuGlnLys 181
Db 910 GCTGTTGTTATTGTAGAGTAATACATCTAGGAAACACGAGTACCTCAATCTCTTAAT 969
Qy 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
Db 970 GAAACCGGACAGTAACCATTCGCGATAGCCGTACAAAGATTATACCTCAAGTAGCGAAA 1029
Qy 202 GlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
Db 1030 GAAGCAGATATTCTTGTGTAGCAGCGCTAGCGAAATTTGTGAAAGAAAGAACTATATC 1089
Qy 222 LysProGlyAlaAlaIleAlaIleAspValGlyIleAsnProValAspAspProGluSerPro 241
Db 1090 AAACCGAGTGGCGTGTGTTATTGATGTTGCGATGATCTCGCGGAAAAAT----- 1137

QY 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
 Db 1138 -----AATAGTATGCGGTGATGGACTTTGATGATGTGTAGAGAGCGGATTT 1191
 QY 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
 Db 1192 ATCAGCGCGGTACCAGGTGGCGGTGCGCGATGACAAATCACAATGCTACTTCCGACACA 1251
 QY 282 LeuGluSerAlaLysArgIleHisLys 290
 Db 1252 TTAAGAGCAGCAAAACGCAATTCGAAA 1278
 RESULT 15
 ABN66178
 ID ASN66178 standard; DNA; 858 BP.
 XX AC ABN66178;
 XX DT 01-JUL-2002 (first entry)
 XX DE Streptococcus polynucleotide SEQ ID NO 269.
 XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX OS Streptococcus pyogenes.
 XX PN W0200234771-A2.
 XX PD 02-MAY-2002.
 XX PF 29-OCT-2001; 2001WO-GB0004789.
 XX PR 27-OCT-2000; 2000GB-00026333.
 XX PR 24-NOV-2000; 2000GB-00028727.
 XX PR 07-MAR-2001; 2001GB-00005640.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;
 PI Tettelin H;
 XX WI; 2002-352536/38.
 XX DR P-PSDB; ABP25547.
 XX PT New Streptococcus protein for the treatment or prevention of infection or
 PT disease caused by Streptococcus bacteria, such as meningitis, and for
 PT detecting a compound that binds to the protein.
 XX FS Claim 7; Page 3181-3182; 4525pp; English.
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus (Streptococcus agalactiae) or group A streptococcus (GAS
 CC Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins
 XX SQ Sequence 858 BP; 279 A; 153 C; 189 G; 237 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,48e-72 Length: 858
 Score: 743.50 Matches: 149
 Percent Similarity: 70.34% Conservative: 55
 Best Local Similarity: 51.38% Mismatches: 77
 Query Match: 50.03% Indels: 9
 Gaps: 4
 DB: 6
 US-10-723-061-18 (1-292) x ABN66178 (1-858)
 QY 2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
 Db 13 GAACATAATTGATGGTAAAGCCTTAGCTCAAAAGATGCAACAAGAGTTAGCAGCTAAAGTC 72
 QY 22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
 Db 73 AACAACTCTAAACAAAAAAGAAATTGTACAGGCTTAGCCGTTATCTTGTAGGTGAT 132
 QY 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGlyIle 61
 Db 133 GATCCTGCTAGTCAGGTGATGTCGGTAATAAGAGCGTCAGCTCTTACTGTAGTTTT 192
 QY 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
 Db 193 AAAAGTGAGACCGTTAGATTATCAGAAATTCATTGTCAAGAGAGCTTATTGCGAGTAATC 252
 QY 82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
 Db 253 GAAGCTTACATGACAGATAACACTATTTCGTGTTATTAGTCAGTACCCCTGCCAAT 312
 QY 102 HisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
 Db 313 CATATTAATGATAAAAAATTAATTCGCCATTGATCCCAAAAAAGATGTGGATGTTTT 372
 QY 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCys 141
 Db 373 CACCGATGAATACAGGTACCTT---TGTCAGGAGCT---CCCTTGATGTTCTTGT 426
 QY 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
 Db 427 ACTCCATCAGGGATTATGGAATTGCTTCGAGATAATAATGTTAACTTGAAGGTAAACAT 486
 QY 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLys 181
 Db 487 GCCGTCAATTATGGCAGATCGAATATCGTTGGAACCAATGCGACAGCTCTTACTGGAC 546
 QY 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
 Db 547 AAAAATCAACAGTCAGCTTGACACATTCAAGAACACGTCATTAGAGAAGATGTGCGC 606
 QY 202 GlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
 Db 607 TGTCCAGATGTTGATTGTGGCAATGCGCAATGCGCAAGGTCATTTCATAACAAACAAATATA 666
 QY 222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSerPro 241
 Db 667 AAAGATGTCGATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 717
 QY 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
 Db 718 -----AAGCTAATTGGAGATGTGGCCTTTTGTAGGTGGCAGAGTTGACGCGAAA 768
 QY 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
 Db 769 ATCACCCTGTACAGAGAGGTGCGGTCTTATGACGATTGCTATGTTGTAGAGCAAACT 828
 QY 282 LeuGluSerAlaLysArg---IleHisLys 290
 Db 829 TATCAATCTGCTCCCGTAGTACTATAAA 858

Search completed: July 24, 2004, 07:19:50

Job time : 625 secs



This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:52:51 ; Search time 91 seconds
(without alignments)
1780.722 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIIGDKLVAKQIREIAVE.....TIAMLLNTLESAKRIHKFK 292

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO.spool/p/US10723061/runat.21072004.090401.7143/app_query.fasta_1.455
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=5.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15
-USER=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-MODE=US10723061@cgn_1_128@runat.21072004.090401.7143 -NCPU=6 -ICPU=3
-NO WMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1030	69.3	926	US-09-903-814A-19	Sequence 19, Appl
2	1016	68.4	1076	US-09-903-814A-15	Sequence 15, Appl
3	740.5	49.8	875	US-09-903-814A-21	Sequence 21, Appl
4	735	49.5	3112	US-09-318-448-9	Sequence 9, Appl
5	724	48.7	13440	US-08-963-527-128	Sequence 128, Appl
6	712.5	47.9	855	US-09-107-532A-3634	Sequence 3634, Ap
7	684	46.0	957	US-09-489-039A-6086	Sequence 6086, Ap
8	681.5	45.9	1830121	US-09-557-884-1	Sequence 1, Appli
9	681.5	45.9	1830121	US-09-643-990A-1	Sequence 1, Appli
10	679.5	45.7	873	US-09-134-001C-564	Sequence 564, App
11	677	45.6	957	US-09-252-991A-13712	Sequence 13712, A
12	677	45.6	1188	US-09-252-991A-13926	Sequence 13926, A

ALIGNMENTS

RESULT 1

US-09-903-814A-19
; Sequence 19, Application US/09903814A
; Patent No. 6680428

; GENERAL INFORMATION:
; APPLICANT: Falco, Carl

; APPLICANT: Ramodu, Lavo O.

; APPLICANT: Orozco, Buddy

; APPLICANT: Rafalski, Antoni

; APPLICANT: Thorpe, Cathy

; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes

; FILE REFERENCE: B81179 USDIV

; CURRENT APPLICATION NUMBER: US/09/903,814A

; PRIOR FILING DATE: 2001-07-12

; PRIOR APPLICATION NUMBER: US 60/092,869

; PRIOR FILING DATE: 1998-07-15

; PRIOR APPLICATION NUMBER: US 09/351,703

; PRIOR FILING DATE: 1999-07-12

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 19

; LENGTH: 926

; TYPE: DNA

; ORGANISM: Glycine max

US-09-903-814A-19

Alignment Scores:

Pred. No.: 2,42e+119 Length: 926
Score: 1030.00 Matches: 194
Percent Similarity: 89.63% Conservative: 22
Best Local Similarity: 80.50% Mismatches: 25
Query Match: 69.31% Indels: 0
DB: 4 Gaps: 0

US-10-723-061-18 (1-292) x US-09-903-814A-19 (1-926)

13 670.5 45.1 9062 4 US-08-956-1715-85 Sequence 85, Appl
14 663.5 44.7 840 4 US-09-134-000C-460 Sequence 460, App
15 661 44.5 873 4 US-09-543-681A-2906 Sequence 2906, Ap
16 638 42.9 1371 4 US-08-958-207A-89 Sequence 89, Appl
17 608.5 40.9 1945 4 US-09-328-352-1366 Sequence 1366, Ap
18 590 39.7 1230025 4 US-09-198-452A-1 Sequence 1, Appli
19 589 39.6 640681 4 US-09-790-988-1 Sequence 1, Appli
20 584.5 39.3 4403765 3 US-09-103-840A-2 Sequence 2, Appli
21 584.5 39.3 4411529 3 US-09-103-840A-1 Sequence 1, Appli
22 544.5 36.6 579 4 US-09-903-814A-17 Sequence 17, Appl
23 541 36.4 956 4 US-09-221-017B-484 Sequence 484, App
24 464 31.2 600 4 US-09-252-991A-13432 Sequence 13432, A
25 385 25.9 669 4 US-09-252-991A-13532 Sequence 13532, A
26 356.5 24.0 578 3 US-09-328-111-765 Sequence 765, App
27 355 23.9 546 4 US-09-252-991A-13632 Sequence 13632, A
28 296 19.9 580073 4 US-08-545-528D-1 Sequence 1, Appli
29 229.5 15.4 4258 4 US-09-596-002-5 Sequence 5, Appli
30 213 14.3 408 4 US-09-252-991A-13433 Sequence 13433, A
31 211 14.2 338 4 US-08-651-155B-33 Sequence 33, Appl
32 211 14.2 338 4 US-09-194-036B-33 Sequence 33, Appl
33 198 13.3 429 4 US-09-596-002-1 Sequence 1, Appli
34 104.5 7.0 1428 4 US-09-252-991A-2357 Sequence 2357, Ap
35 104.5 7.0 1449 4 US-09-252-991A-2359 Sequence 2359, Ap
36 104.5 7.0 2343 4 US-09-252-991A-2275 Sequence 2275, Ap
37 103 6.9 9280 4 US-08-956-171B-131 Sequence 131, App
38 100 6.7 246 4 US-09-543-681A-2795 Sequence 2795, Ap
39 99.5 6.7 696 4 US-08-426-630-53 Sequence 53, Appl
40 99.5 6.7 955 4 US-08-426-630-52 Sequence 52, Appl
41 98 6.6 1149 4 US-09-134-001C-928 Sequence 928, App
42 96 6.5 1664976 4 US-08-916-421B-1 Sequence 1, Appli
43 92 6.2 3078 4 US-09-711-164-272 Sequence 272, App
44 90.5 6.1 1230 4 US-09-134-001C-953 Sequence 953, App
45 89 6.0 18475 4 US-08-961-527-38 Sequence 38, Appl


```
Db      872 ACGTGGATGGGCAAGCG 892
RESULT 3
US-09-903-814A-21
; Sequence 21, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famedu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 21
; LENGTH: 875
; TYPE: DNA
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (584)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (680)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (699)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (704)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (756)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (767)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (771)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (780)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (800)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (802)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (816)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (834)

; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (837)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (842)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (846)..(847)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (858)..(859)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (870)
; OTHER INFORMATION: n is a, c, g or t
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (874)
; OTHER INFORMATION: n is a, c, g or t
; OTHER INFORMATION: n is a, c, g or t
US-09-903-814A-21

Alignment Scores:
Pred. No.:      4,35e-83      Length:      875
Score:          740.50      Matches:    146
Percent Similarity: 79.31%      Conservative: 38
Best Local Similarity: 62.93%      Mismatches: 47
Query Match:    49.83%      Indels:     2
DB:              4              Gaps:       0

US-10-723-061-18 (1-292) x US-09-903-814A-21 (1-875)

Qy      1 AlaLysIleAlaAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      20 GGCAAATCATCGACGGCAAGCCCATCGCCGCCGAAATCAGCGCGAGATCGGCGCGAG 79
Qy      21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      80 GTCCCGCTCTCTCGTCCGCCCCACACATCGTGGCGGGCTGGCGTGTGTGTCGTGGGG 139
Qy      41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      140 AGCAGGAGGACTCGCAGACGTACGTGCAGATGAAGCGCAAGGCCCTGCCCGAGGTCCGC 199
Qy      61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      200 ATCCGCTCTTTCAGCTCGACTCCCGGAGACATCGCCGAGCGCGCGCTGTCGCGCGAG 259
Qy      81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      260 GTCCACCGGCTCAACGCCGACCGCGCGCTCCACCGGAATCTTTGTTTCAGCTTCATTGCC 319
Qy      101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      320 AAGCATATCAAGAGAAATATCTTAAACCAAGATCTCCATTGAGAAAGATGTCGACGCG 379
Qy      121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      380 TTTCATCCCTTGAACATTGGCAAGCTTGCAATGAAAGGCAGAGATCCACTGTTGCTACCT 439
Qy      141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      440 TGCACGCCAAAGGATGTCATGAGTCTCTGTCCGAAAGTGGCGCTCACTATAAAGAAAA 499
Qy      161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAla-AlaLeuLeuLeuGlu 180
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      500 CAGCGAGTTGTGGTGGGGGTAGCAACATCGTGGGTTTACCAAGATATCCTCTCTCTCT 559
```

```

180 nlyslalasnalanthrValSerlleValHisSerAsnThrLysLysProGluGluIlelh 200
Db 560 GAAAGCGGACGCTACCGTGTCATNGTCATCAACGGACCCAAAT-CCCCAACAAATTC 618
Qy 200 rargGlnAlaAspIleValIleAlaValAlaValGlyValAlaAsnLeuValArgGlyserTr 220
Db 619 CGTCAAGCAAGACATTGTTCATTGCGAGCGAGCTGGGCAAGCCATGATCATCAAGGAGACTG 678
Qy 220 pileyLysProGlyAlaAlaIlelleAspValGly 231
Db 679 GNTTAAACAAAGCGCAACGCNCAATCNACGTCGGG 712

RESULT 4
US-09-318-448-9
; Sequence 9, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Scenocos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3112
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-9

Alignment Scores:
Pred. No.: 1,676-81 Length: 3112
Score: 735.00 Matches: 145
Percent Similarity: 67.91% Conservatise: 56
Best Local Similarity: 48.99% Mismatches: 91
Query Match: 49.46% Indels: 4
DB: 3 Gaps: 3

```

```

RESULT 5
US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-128
Alignment Scores:
Pred. No.: 4.3e-79 Length: 13440

```

Score: 724.00 Matches: 143
Percent Similarity: 68.8% Conservative: 54
Best Local Similarity: 50.0% Mismatches: 81
Query Match: 48.72% Indels: 8
DB: 4 Gaps: 2

US-10-723-061-18 (1-292) x US-08-961-527-128 (1-13440)

QY 2 LysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIle 21
DB 1558 CAGATTATGATGGGAAGCTTTACGGCCCAATTCAGGGCGAGTGGCTGAAAGACT 1617

QY 22 AlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySer 41
DB 1618 GCAGAAATTAAGGAAGAACAGCTAGTGGCTGGTTGGTAGTCATTTTGGTTGGGAC 1677

QY 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIle 61
DB 1678 AATCAGCAGCCAGCAAGTCTAGTTCGCAACAGGAGAGGTGAGCCCTTCGGCGGTTC 1737

QY 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
DB 1738 CGTAGCGAAGTAGTAGCGGTTCAGAGACCATTAATCAAGAGGAATTTGTTAGACCTGATT 1797

QY 82 AlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHis 101
DB 1798 GCTAATAACAATCAGATCCAGCTTGGCATGGATTTTGGTTGAGTTGCCATTACCAAAA 1857

QY 102 HisMetAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
DB 1858 CACATTGATGAAGAGCGGTCTATTGGCTATTGACCCAGAAAGGATGTGATGGTTTC 1917

QY 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCys 141
DB 1918 CATCTCTAAACATGGGCGCTT-----TGGTCTGCTCATCCAGTCATGATCCTTCG 1971

QY 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
DB 1972 ACACCGCAGGAATATGGAATGTTCCATGAATATGGATGACTTGAAGGTAAAAAT 2031

QY 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuLys 181
DB 2032 GCAGTCGTCATCGGTCGATCCAAATATTGCGAAACCTATGGCCAGCTCTTTTGGCA 2091

QY 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArg 201
DB 2092 AAGATGCAACAGTAACCTTGAATCTCACTCACTCATCTATATCTTCCAGGTGGCTGCA 2151

QY 202 GlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIle 221
DB 2152 AAAGCAGATATTCTGGTTGTTGCAATCGTGTGCGCAAGTTTGTGCTGCTGACTTGTG 2211

QY 222 LysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSerPro 241
DB 2212 AAACAGGTGGGTAGTCAATGACCTGGATGAAACCGCGATGAAATGGT----- 2262

QY 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGlyGluAlaSerLysIleAlaGlyLeu 261
DB 2263 -----AAGCTCTGGGGATGTTGATATGATGGCGGTGGCCCACTTGTAGCCAC 2313

QY 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr 281
DB 2314 ATTACGCCAGTCCCTGGAGGTTCGCTCTATGACCATTAATCTATGCTGATGGAGCAAAC 2373

QY 282 LeuGluSerAlaLysArg 287
DB 2374 TATCAGGCGACCTTAGG 2391

RESULT 6
US-09-107-532A-3634
; Sequence 3634, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3634:
SEQUENCE CHARACTERISTICS:
LENGTH: 855 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...855
SEQUENCE DESCRIPTION: SEQ ID NO: 3634:
US-09-107-532A-3634

Alignment Scores:
Pred. No.: 1,36e-79 Length: 855
Score: 712.50 Matches: 146
Percent Similarity: 69.93% Conservative: 54
Best Local Similarity: 51.05% Mismatches: 77
Query Match: 47.95% Indels: 9
DB: 4 Gaps: 5

US-10-723-061-18 (1-292) x US-09-107-532A-3634 (1-855)

QY 1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
DB 4 GCAGATTGATTAACGAAAGAAATAGCTGAAAAATCGAGGACAGAGATTGCTGAAAA 63

QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
DB 64 ATCCAAAAATTGAAAGATAAT---GGTATCCATCCCGGATTAGTTGTTTATTGTTAGG 120

QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
DB 121 GAAAAATCCAGCCAGCCAAATCTATGTAGAAATAAAGAACGCTAGCAAAAGAAATCGGA 180

QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
DB 181 ATCATTCATTGGTAGAGCGTTATCTCTGACAGCAITTCGAAAGACGACTACTTCGAGAA 240


```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1830121 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.: 2,68e-70 Length: 1830121
Score: 681.50 Matches: 144
Percent Similarity: 66.67% Conservatives: 44
Best Local Similarity: 51.06% Mismatches: 85
Query Match: 45.86% Indels: 9
DB: 4 Gaps: 2

US-10-723-061-18 (1-292) x US-09-643-990A-1 (1-1830121)
QY 1 AlaLysIlelleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 640319 GCAAAATCATCTCTGCTACTGAACTTCAAAACAAATTAAGCTAATTTAGCAGATAA 640260
QY 21 IleAlaLysMetLysAspAlaIlelleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 640259 ATAACCCATATATTGACCAAGAAACCGCTCCAGGACTTCTGTATATCTCTGTTGT 640200
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 640199 GCTGATCTCTGCTTCCGCAAAATTTATGTGGCAATAAAGCTAAAGTTGCGAGGAAGTCGT 640140
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLysHis 80
Db 640139 ATTCTCTCTAAATCTTATGACCTCCCTGAAACCAACCCCAAAATGAATTTAGCGATT 640080
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 640079 ATTGACCAACTTAACGCAGATAAAATATTGCGGTATTTTGTGCAACTCCCTTGCCA 640020
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 640019 AAACAAATTAATCGGAAGCTATTTATGAACTGATCGATCTTAAAGATGATGATGCG 639960
QY 121 PheHisProLeuAsnIlelleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 639959 TTTCATCTTATATGTTGGCGGTTATGCCAA-----CGTATTCCACTTTACGTGCT 639906
QY 141 CysThrProLysCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 639905 TGTACACCTTATGTTGATGAACACTATTAGAACCAACCGGTATTGATTACACGGCAAG 639846
QY 161 ArgAlaValIlelleGlyArgSerAsnIlelleValGlyMetProAlaAlaLeuLeuGln 180
Db 639845 CACGAGTAATTTAGCGCATCAATATTTGCGCGCCCAATGCTTTTGAATTTA 639786
QY 181 LysAlaAsnAlaThrValSerIlelleValHisSerAsnThrLysLysProGluLeuThr 200
Db 639785 TTAGCAGGGGCAACCGCTCAGATACATATGATTGCAATATTCACCTTAAATAATTTGGAACACCGCTT 639726
QY 201 ArgGluAlaAspIlelleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db 639725 CGCCAGCAGATATCTTGTGTTGGTAGGTTAAACCAAAATTTAATTTCTGCGCATTTG 639666
QY 221 IleLysProGlyAlaAlaIlelleAspValGlyIleAsnProValAspAspProGluSer 240
Db 639665 ATCAAGAAAGTCGGTAGTATGATTGTTGAAATTAACCGAGTAGATGGA----- 639615
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluLysSerLysIleAlaGly 260
Db 639614 -----AAATAGTAGCGGATGATGATTTGATTAACCGAGTAGATGGA----- 639567
QY 261 LeuIleThrProValProGlyValGlyValProMetThrIleAlaMetLeuLeuSerAsn 280

```

```

Db 639566 TACATTACACAGTCCCGGTGGCGTGGCCAAATGACCGTTCGAATGTTGATGAGCAAT 639507
QY 281 ThrLeu 282
Db 639506 ACGCTT 639501

RESULT 10
US-09-134-001C-564
; Sequence 564, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: LYTH Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 564
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-564

Alignment Scores:
Pred. No.: 1.92e-75 Length: 873
Score: 679.50 Matches: 139
Percent Similarity: 69.23% Conservatives: 59
Best Local Similarity: 48.60% Mismatches: 79
Query Match: 45.73% Indels: 9
DB: 4 Gaps: 3

US-10-723-061-18 (1-292) x US-09-134-001C-564 (1-873)
QY 1 AlaLysIlelleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 19 GCTAAATTTTATGATGTTAACTAACTAGCAAGATATAGCAAGACTTAAATCAAA 78
QY 21 IleAlaLysMetLysAspAlaIlelleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 79 GTCAATGACTTAAAGAA--TATGGTTTCTCCAAATATATCATCATATATTAGTTGTT 135
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 136 AATGATGTCAGCCCAAGCTTATGTAATCAAGAAATCAAGAAACACGACAGAAAATTTGA 195
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLysHis 80
Db 196 ATGATATCAGAAATTTATCTACTTAGACGAATCCACATCAGACGAAGTAGTTTAAAGCAA 255
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 256 CTTAATCGATTAAACAAATGACGATCTGTTAGTGGTATTTTACTTCAAGTACCATTACCT 315
QY 101 HisHisMetAsnAspGluAsnIlelleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 316 AACAGGTAAACGACCAAAAGTACTTTGAAGCTATCAATCCGGAAGAGATGTAGATGTT 375
QY 121 PheHisProLeuAsnIlelleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 376 TTCCATCCAAATTAATTTGAAATTTGATATATATGAT-----GAGCAAACTTTGATCA 429
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLysGlyLys 160
Db 430 TGTACACCTCTGGTATTATGGAATTTTAAACACGACGATATTAATTAGAGGAAA 489
QY 161 ArgAlaValIlelleGlyArgSerAsnIlelleValGlyMetProAlaAlaLeuLeuGln 180
Db 490 AATGCACTTGTATTGTTGCTGATCATATTGTTGGCAACCTGTTTCAAGATTTATTGTTG 549

```


[illegible]

```

RESULT 11
US-09-252-991A-13712
; Sequence 13712, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOTUBERCULOSIS
; TITLE OF INVENTION: ASCLINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252-991A

```

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13712
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13712

```

Alignment Scores:	
Pred. No.:	4,598-75
Score:	577.00
Percent Similarity:	64.36%
Best Local Similarity:	49.83%
Query Match:	45.56%
DB:	4
Length:	957
Matches:	144
Conservative:	42
Mismatches:	95
Indels:	8
Gaps:	2

US-10-723-061-18 (1-292) x US-09-252-991A-13712 (1-957)

Qy 1 AAlalysIlelleAspGlyLysLeuValAAlalysGlnIlelArgGluCluIleAAlaValGlu 20
|||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 109 GCAAACTCATCGCGGGAAGCGATCGCGCCAACTTCGCCAGAGATGCCCAACGC 168

Qy 21 ILeAlaLysMetIysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
::: ::: :::
Db 169 GTGACGAGCGGCGGCAGCAAGGCGTGGCGCTTCCGCGCTGGCGGTGATCCTGGTGGC 228

Qy 41 SerArgLysAspSerGlnThrTrpValArgAsnLysLysLysAlaCysGluAlaValGly 60
:: :: :: :: :: :: ::
Db 229 ACGAGTCGGCCCTCTCAGATCATGTGGCGCACAAAGCGCAGGACTCGCAGGAAGTCGC 288

QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
289 TTTCTCTCCAGGCTTACGATTCTCCCGCAAAACCCAGGACGACCTCTGGGCGCTG 348

Qy	81	IleAlaThrPheAsnSerAspProSerValHisGlyIleuValGlnIleuProLeuPro	100
Db	349	ATCGACCGCCTGAACGACGACCCCGCATCGCGGCATCTGTGCAGCTACCCCTGCCCC	408
Qy	101	HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly	120
Db	409	GCCACCTGGACGCTCCCTGCTGCGAGCGCATCCACCGCAGAAGACGTCGACGCT	468
Qy	121	PheHisProLeuAsnIleGlyArgLeuAlaMetGlnIlyArgAspProPhePheValPro	140
Db	469	TTCATCCCTACACATCGCGCGCTGGCCAG-----CGATGCGCTCTCGCGCCCC	522
Qy	141	CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys	160
Db	523	TGCACCCCGAAAGGATCATGACCTGCTGCCAGCACCGCGCCGACCTGTACGCGCATG	582
Qy	161	ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaIleLeuLeuGln	180
Db	583	GATCGGTGCTGTGGCGCCTCGAACATCGTCGCGCGGCCCATGCTGTGGAGTGTCTG	642
Qy	181	LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr	200
Db	643	CTGGGTGGCTGACCGCTCACCGTCACCCACCGCTTCACCGGCACCTGGCGACCATGTG	702
Qy	201	ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAlaAsnLeuValArgGlySerTrp	220
Db	703	TCGCGCGCGACCTGTGTGTGCTGCCGCAAGCCGGGACTGTGTCAAGGCGAGTGTG	762
Qy	221	IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer	240
Db	763	ATCAGGAAGGCGCATCGTATGACGTGCGCATCAACCGCACCGCGCGACGGC-----	816
Qy	241	ProArgGlyTyrArgLeuValGlyAspValCysTyrGluAlaSerIlyIleAlaGly	260
Db	817	-----CGCCTGCTCGGCGACGTGGAAATACGAAGTGCAGCGCAACCGCGCAGC	864
Qy	261	LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn	280
Db	865	TGATCACCCTCCCGTCCGGCGGGCTGTGGGCCGATGACCCGCGCTGCTCTGGAAAAAT	924
Qy	281	ThrLeuGluSerAlaLysArgIleHis	289
Db	925	ACCTTCACGCGCGCGAACACTGCAC	951

RESULT 12
US-09-252-991A-13926
; Sequence 13926; Application US/09252991A

```

; PATIENT NO. 6551795
;
; GENERAL INFORMATION:
;
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMIN
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNO
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13926
; LENGTH: 1188
;

```

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13926

```

Alignment Scores:	6.53e-75	Length:	1188
Pred. No.:	677.00	Matches:	144
Score:	64.3%	Conservative:	42
Percent Similarity:	49.8%	Mismatches:	95
Best Local Similarity:	45.5%	Matches:	8
Query Match:	4	Indels:	2
DB:		Gaps:	2

US-10-723-061-18 (1-292) x US-09-252-991A-13926 (1-1188)

Qy 1 AlaLysIleLeuAspGlyLysLeuValAlaLysGlnIleArgGluGluLeuAlaValGlu 20
Db 260 GCACAACTGATCGACGCGCAAGGATCGCGCCACCTTCGCGCAGGATACCCCAAGC 319
Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 320 GTACCGGAGCGCGCGCAAGGCTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTCGCGC 379
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
Db 380 ACCGATCCGCGCTTCAGGCTATGTGGCGCACAGGCAAGGATCGCGAAGTTCGCGC 439
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 440 TTTCTCTCCAGCGCTACGATCTTCGCGCGCAAGCAAGCAGGACGACCTTCGCGCGCTG 499
Qy 81 IleAlaThrPheAsnSerSerProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 500 ATCCAGCGCTGACGACCGCGCCATCGCGCATCTCTGTCAGCTACCCCTGCGCC 559
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 560 GCCCACCTGAGCGCTCCCTGCTGCTGGAGCGCATCCACCGCGCAAGGACGTCGACGCT 619
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 620 TTCATCCCTTACAACTCGCGCGCTGCGCCAG-----CGATGCGCTCTCTGCGCGCC 673
Qy 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 674 TGCAACCGCGAAGGCATCATGACCTGCTCGCGACCGCGCGCGCGCGCGCGCGCGCATG 733
Qy 161 ArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 734 GATCGCGTGTGTCGCGCGCTCGAACATCGTCGCGCGCGCGCGCGCGCGCGCGCGCGT 793
Qy 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 794 CTGGTGTGTGACCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
Qy 201 ArgGluAlaAspIleValIleAlaValAlaValAlaAsnLeuValArgGlySerTrp 220
Db 854 TCAGCG 913
Qy 221 IleLysProGlyAlaAlaIleLeuAspValGlyIleAsnProValAspProGluSer 240
Db 914 ATCAGAGAGCG 967
Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 968 -----CGCGTGTGTCG 1015
Qy 261 LeuIleThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 1016 TGGATCACCG 1075
Qy 281 ThrLeuGluSerAlaLysArgIleHis 289
Db 1076 ACCCTGCG 1102

RESULT 13
US-08-956-171E-85
; Sequence 85, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash

Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 45,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 9062 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 85:
US-08-956-171E-85

Alignment Scores:
Pred. No.: 1,16e-72 Length: 9062
Score: 670.50 Matches: 139
Percent Similarity: 68.88% Conservative: 58
Best Local Similarity: 48.60% Mismatches: 80
Query Match: 45.12% Indels: 9
DB: 4 Gaps: 3

US-10-723-061-18 (1-292) x US-08-956-171E-85 (1-9062)

Qy 1 AlaLysIleLeuAspGlyLysLeuValAlaLysGlnIleArgGluGluLeuAlaValGlu 20
Db 6676 GCTAAATTTTAGATGTTAAACAAATTCGCAAGACTACAGACGGGTACAGATCAA 6735
Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 6736 GTTGAACCGCTAAAGAAAG---GGTTTACCTTAATATCCGTTATATTAGTTGTT 6792
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
Db 6793 AATGATCGCGCTAGTCAAGATTATGTTAGTCAAAAAGAAAGCAGCTGAAAAAATTGTT 6852
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 6853 ATGATTTCAAGAAATCGTACATTTTGGAGAAACAGCTACTGAAAGAAAGATTATTAACGAA 6912
Qy 81 IleAlaThrPheAsnSerSerProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 6913 CTAATAGACTAAATAATGATGATTTCTGTAAGTGGTATTTTGTTCAAGTACCATTAACCA 6972
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 6973 AAACAAGTTAGCGAAGACAGAAATATTAGAACCAATCAATCTCTGAAAAAGATGTGGACGTT 7032

QY 121 PheHisProLeuAsnIleGlyValGluAlaMetGlnGlyArgAspProPheValPro 140
Db 7033 TTTTATCCAAATAATATAGGAATATATATCATGAT-----GAACAACTTTGTACTT 7086
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 7087 TGCACACCGCTCGGCATCATGGAATATATAAACATGCTGATATGATTTAGAGGTAAA 7146
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 7147 AATGCAGTTGTATGTGACGAGTCATATGTCGACACACAGTTTCTAAGTACTACTT 7206
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 7207 CAAAAAATGATCATGATCAATCTTACATCTCGTTCCAAAGATATGGCATCATATTTA 7266
QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAlaAsnLeuValArgGlySerTrp 220
Db 7267 AAAAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7326
QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 7327 GTCAAAGAGGAGCAGTAAATATATGATGATGATGATGATGATGATGATGATGATGAT 7380
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 7381 -----AAATTAAGGTGACGTTGATGATGATGATGATGATGATGATGATGATGAT 7428
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 7429 GCTATTACACAGTTCCTGCTGGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7488
QY 281 ThrLeuGluSerAlaLys 286
Db 7489 ACTTTGCTTGCAGAAAAA 7506

RESULT 14

US-09-134-000C-460

; Sequence 460, Application US/09134000C

; Patent No. 6617156

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; PRIOR FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 460

; LENGTH: 840

; TYPE: DNA

; ORGANISM: Enterococcus faecalis

US-09-134-000C-460

Alignment Scores:

Pred. No.:	1,83e-73	Length:	840
Score:	663.50	Matches:	135
Percent Similarity:	68.52%	Conservative:	50
Best Local Similarity:	50.00%	Mismatches:	76
Query Match:	44.65%	Indels:	9
DB:	4	Gaps:	5

US-10-723-061-18 (1-292) x US-09-134-000C-460 (1-840)

QY 3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
Db 34 GTAATTACGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 93
QY 23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42

Db 94 AAGATGACACAAACAA-----GGCATCCAACCCAGGATTAGTTGTTTTATTAGTTGGGAAAT 150
QY 43 LysAspSerGluThrTyrValArgAsnLysLysAlaCysGluAlaValGlyIleLys 62
Db 151 CCTGCCAGCAACAGTATGTGAGAAATAAGAACGTCGACAGCCAAATTCGCAATTCG 210
QY 63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
Db 211 TCAAGAGTCGAAAACTGCCAGAACTATTTCAGAGAAGAATATTGGCTGAGATTGAC 270
QY 83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
Db 271 AATATAATCAAGATTCACGCTTTATGGCANTCTTGTACAACTACCTTTGCCAAACAT 330
QY 103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
Db 331 ATTGATGAAGAAAAAGATTTTATTAGCGATTGATCCAAAAAAGAGTGTAGATGGTTCCAT 390
QY 123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCysThr 142
Db 391 CCAATGAATTTAGCCGTTTGTGTA---GGCAAA---CCTGAATGATTTCTTTCACG 444
QY 143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAla 162
Db 445 CCGTATGGAATTATGAAGATGTTTGAAGCTTATGATATTGATTAAACAGTAAACGTCG 504
QY 163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLysAla 182
Db 505 GTGGTTATTGTCGAAAGCAATATTGTTGCAAGCCCAATGCGCAATATTGTTAATGAA 564
QY 183 AspAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
Db 565 AATGCAAGCTGACGATGCCCATTTCTAAACTGAACATTTAGCAGAGTTGCCAAAGAA 624
QY 203 AlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
Db 625 GCGGATATTTTAGTTAGCAATTGACGCGGCGCATTTTGTCCACCAAGAGTTTCGTGAA 684
QY 223 ProGlyAlaAlaIleIleAspValGlyIleAsnProValAspProGluSerProArg 242
Db 685 CCAGGTCGGTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 729
QY 243 GlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeuIle 262
Db 730 GGA---AAACTGATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 786
QY 263 ThrProValProGlyValGlyProMet 272
Db 787 ACGCCGGTTCCCAAGGTGTTGCCCAATG 816

RESULT 15

US-09-543-681A-2906

; Sequence 2906, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 2906

; LENGTH: 873

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-09-543-681A-2906

Alignment Scores:

Pred. No.:	4.01e-73	Length:	873
------------	----------	---------	-----

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 08:50:33 ; Search time 575 Seconds

(without alignments)
2482.406 Million cell updates/sec

Title:
Perfect score: 1486

Sequence: 1 AKIDGKLVAKQREEIAVE.....TIAMLLSNTLESAKRIHKK 292

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3216467 seqs, 2444149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090403_7226/app_query.fasta_1.455
-DB=Published Applications NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10723061@cgn2_1_912@runat_21072004_090403_7226
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2.*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Query Match Length DB ID Description

ALIGNMENTS

US-09-938-842A-541

; Sequence 541, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 541
; LENGTH: 1083

1 1238 83.3 1083 9 US-09-938-842A-541 Sequence 541, App
2 1238 83.3 1083 11 US-09-938-842A-541 Sequence 541, App
3 1234 83.0 1632 13 US-10-424-599-91222 Sequence 19, Appl
4 1030 69.3 926 9 US-09-903-814-19 Sequence 19, Appl
5 1020 69.3 926 17 US-10-723-061-19 Sequence 33029, A
6 1025 69.0 1066 13 US-10-425-114-33029 Sequence 23607, A
7 1025 69.0 1066 13 US-10-425-114-33029 Sequence 23607, A
8 1019 68.6 1798 13 US-10-424-599-11288 Sequence 11296, A
9 1019 68.6 1798 13 US-10-424-599-11288 Sequence 11296, A
10 1016 68.4 1076 9 US-09-903-814-15 Sequence 15, Appl
11 1016 68.4 1076 17 US-10-723-061-15 Sequence 15, Appl
12 1012 68.1 1356 17 US-10-437-963-74482 Sequence 74482, A
13 1010 68.0 1286 13 US-10-424-599-125693 Sequence 125693, A
14 994 66.9 1409 13 US-10-425-114-9450 Sequence 9450, Ap
15 988 66.5 1413 17 US-10-437-963-23018 Sequence 23018, A
16 965 64.9 1561 13 US-10-425-114-28648 Sequence 28649, A
17 919 61.8 2442 17 US-10-437-963-52595 Sequence 52595, A
18 801 53.9 837 13 US-10-425-114-2899 Sequence 2899, Ap
19 775 52.2 837 16 US-10-369-493-33838 Sequence 33838, A
20 763.5 51.4 495269 16 US-10-398-221-8 Sequence 8, Appl
21 763.5 51.4 301208 16 US-10-398-221-2058 Sequence 2058, Ap
22 751.5 50.6 5635 16 US-10-398-221-3623 Sequence 3623, Ap
23 740.5 49.8 875 9 US-09-903-814-21 Sequence 21, Appl
24 740.5 49.8 875 17 US-10-723-061-21 Sequence 2103, Ap
25 735 49.5 3112 9 US-09-880-107-2103 Sequence 367, App
26 735 49.5 3112 12 US-09-968-007A-367 Sequence 40, Appl
27 735 49.5 3112 15 US-10-171-581-40 Sequence 40551, A
28 730 49.1 1035 16 US-10-363-493-40551 Sequence 238, App
29 724 48.7 13440 13 US-09-787-787-239 Sequence 128, App
30 724 48.7 13440 16 US-10-158-844-128 Sequence 45509, A
31 722.5 48.6 891 16 US-10-369-493-45509 Sequence 41046, A
32 716 48.2 840 16 US-10-369-493-41046 Sequence 35899, A
33 715 48.1 900 16 US-10-369-493-43170 Sequence 43170, A
34 711 47.8 855 16 US-10-369-493-43170 Sequence 42987, A
35 710.5 47.8 840 16 US-10-369-493-43170 Sequence 42987, A
36 707.5 47.6 885 16 US-10-369-493-42077 Sequence 42077, A
37 704.5 47.4 7528 9 US-09-070-927A-55 Sequence 55, Appl
38 694.5 46.7 834 16 US-10-369-493-41608 Sequence 41608, A
39 690.5 46.5 901 16 US-10-369-493-31770 Sequence 31770, A
40 687.5 46.3 852 16 US-10-369-493-34559 Sequence 34559, A
41 687.5 46.3 855 9 US-09-974-300-998 Sequence 998, App
42 684.5 46.1 840 16 US-10-369-493-40237 Sequence 40237, A
43 683.5 46.0 840 16 US-10-369-493-30840 Sequence 30840, A
44 683.5 46.0 852 16 US-10-369-493-46889 Sequence 46889, A
45 681.5 45.9 1830121 15 US-10-329-960-1 Sequence 1, Appl

TYPE: DNA
 ORGANISM: Arabidopsis thaliana
 US-09-938-842A-541

Alignment Scores:

Pred. No.: 1.88e-146 Length: 1083
 Score: 1238.00 Matches: 231
 Percent Similarity: 91.44% Conservative: 36
 Best Local Similarity: 79.11% Mismatches: 25
 Query Match: 83.31% Indels: 0
 DB: 9 Gaps: 0

US-10-723-061-18 (1-292) x US-09-938-842A-541 (1-1083)

```

Qy 1 AlalysilleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 205 GCAATTGTAATAGATGGAAGGCTGTGGCAAAAGATTAGATGAATCAAAATGAA 264
Qy 21 IleAlaLysMetLysAspAlaIleGlyValProGlyLeuAlaValIleLeuValGly 40
Db 265 GTTTCAGAATGAAGGAATCAATGTGTCATCTGTTAGCAGTAATCCTTTGTGGT 324
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAAGACTGTGACTCCGTGGA 384
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 385 ATCAATCGTTTGAAGTTCTGCTAGCTGAAGATTCATCAGAAAGAGGTGTGAATCT 444
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGATTCATATGATGATGCTTCTGTCATGGAATCCTTGTTCAGTTCCTCGCA 504
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGATGATGACAGACATCTGATGCGTTAGTATAGAGAAAGATGTGACGA 564
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 565 TTTTCATCCGCTAAATATGGAAGGCTTGGCATGCGTGAAGAAACCCCTTATTCGTCT 624
Qy 141 CysThrProLysGlyCysMetGluLeuHisArgTyrGlyValGluLysLysLys 160
Db 625 TGTACTCCAAAAGATGCAATGATGTTGTCATAGATACAACTTGAATCAAGAAAG 684
Qy 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGAGCGGTTGTTATCGAAGGAGTAACATGTCGTTATGCCAGTCTCTTTACTGCAG 744
Qy 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluLysLeuThr 200
Db 745 AGGAGGATGCAACCGTTAGCATTTCCATTCAGAACCAAGAACCCCTGAAGAAATCA 804
Qy 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTyr 220
Db 805 AGAAGAGCTGATATTATATCTCAGCTGTGACAGCCCAACATGTCAGAGAAAGCTGG 864
Qy 221 IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 865 ATAAACCGGGCGAGCTCATCATGATGTGGATTAATCCTGTTGAGATCCCAAGTCT 924
Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 925 GCGCGTGGATATCGATTGTTGGAGACATTTCTACGAGGAGGCTAGCAAGTTGCATCA 984
Qy 261 LeuIleThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 985 GCATCACACCTGTTCTCGGCGGTGTAGACCAATGACCATAGCCATGCTCTTATCCAAC 1044
Qy 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1045 ACTTTAAACATCAGCTAAGAGGATTACCAACTTCCAG 1080

```

RESULT 2

US-09-938-842A-541
 ; Sequence 541, Application US/09938842A
 ; Publication No. US20040009476A9

GENERAL INFORMATION:

APPLICANT: Harper, Jeff
 APPLICANT: Kreps, Joel
 APPLICANT: Wang, Xun
 APPLICANT: Zhu, Tong

FILE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
 TITLE OF INVENTION: SAME, AND METHODS OF USE

FILE REFERENCE: SRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

CURRENT FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 541

LENGTH: 1083

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-541

Alignment Scores:

Pred. No.: 1.88e-146 Length: 1083
 Score: 1238.00 Matches: 231
 Percent Similarity: 91.44% Conservative: 36
 Best Local Similarity: 79.11% Mismatches: 25
 Query Match: 83.31% Indels: 0
 DB: 11 Gaps: 0

US-10-723-061-18 (1-292) x US-09-938-842A-541 (1-1083)

```

Qy 1 AlalysilleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 205 GCAATTGTAATAGATGGAAGGCTGTGGCAAAAGATTAGATGAATCAAAATGAA 264
Qy 21 IleAlaLysMetLysAspAlaIleGlyValProGlyLeuAlaValIleLeuValGly 40
Db 265 GTTTCAGAATGAAGGAATCAATGTGTCATCTGTTAGCAGTAATCCTTTGTGGT 324
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 325 GACAGAAAAGATTCTGCAACTTATGTGAGGAACAAGAAAGACTGTGACTCCGTGGA 384
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 385 ATCAATCGTTTGAAGTTCTGCTAGCTGAAGATTCATCAGAAAGAGGTGTGAATCT 444
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGATTCATATGATGATGCTTCTGTCATGGAATCCTTGTTCAGTTCCTCGCA 504
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGATGATGACAGACATCTGATGCGTTAGTATAGAGAAAGATGTGACGA 564
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 565 TTTTCATCCGCTAAATATGGAAGGCTTGGCATGCGTGAAGAAACCCCTTATTCGTCT 624
Qy 141 CysThrProLysGlyCysMetGluLeuHisArgTyrGlyValGluLysLysLys 160
Db 625 TGTACTCCAAAAGATGCAATGATGTTGTCATAGATACAACTTGAATCAAGAAAG 684
Qy 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGAGCGGTTGTTATCGAAGGAGTAACATGTCGTTATGCCAGTCTCTTTACTGCAG 744

```

```
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 745 AGGAGGATGCAACCGTTAGCAATATCCATTCAAGAACCAAGAACCCCTGAAGAAATCACA 804

QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 805 AGAAGCTGATATTAAATCTCAGCTGTTGGACAGCCAAACAGGTGTCAGAGAACTGG 864

QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspProGluSer 240
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 865 ATAAACCGGGCGCAGTCCATCATCGATGTTGGGATTAACTCTGTGAGGATCCAAGTGCT 924

QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 925 GCGCGTGGATATCGATTGGTTGGAGACATTTCCTACGAGGAGGCTAGCAAGTTGTCATCA 984

QY 261 LeuIleThrProValProGlyValGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 985 GCCATCACACCTGTTCTCGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAC 1044

QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1045 ACTTTAATCAGCTAAGAGGATTCAACAACCTCCAG 1080

RESULT 3
US-10-424-599-91222
; Sequence 91222, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalick David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91222
; LENGTH: 1632
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53384C.1
US-10-424-599-91222

Alignment Scores:
Pred. No.: 1,13e-145 Length: 1632
Score: 1234.00 Matches: 233
Percent Similarity: 89.73% Conservative: 29
Best Local Similarity: 79.79% Mismatches: 30
Query Match: 83.04% Indels: 0
D: 13 Gaps: 0

US-10-723-061-18 (1-292) x US-10-424-599-91222 (1-1632)

QY 1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 309 GCTAGGTGATTGATGGAAATCCCTTGCCAGCAAAATCAGAGATGAGATACCGCTGAG 368

QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 369 GTTCCAGGATGAGAAATCCATTCGTTGATTCCTCGGTTGGCTGTAATCTTGTGGG 428

QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 429 GATAGAAAGATCTCAGCACTTATGTGCTACAGAGAGAGGCTTGTGAATCTGTGGA 488

QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 489 ATCAATTCCTTGGAGCAATCTGCCAGAGGATCCACAGAGAAAGATTTTGAACAT 548
```

```
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGluLeuProLeuPro 100
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 549 ATTGAGGCTCAATATGATGATCCCTTCAGTTCATGGCATCCTTGTTCAGTTACCTTACT 608

QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 609 TCGCATATGATGATGAGCAGACATCTTGAACGCTGTTCAGGATTTGAGAGGATGATGAT 668

QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 669 TTTTCATCCGTTAAATATTGCTGCTGCTATGCTGTGGAAGAAACCTCTGTGTTGCTCT 728

QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 729 TGTACACCAAGGATGATAGAGTACTGCACAGATCAATGTTCTTATTAAAGGAAAG 788

QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 789 AGGCGTGTGTTGATGTCGAGCAATATTGTTGGAATGCCAGCTGCTCTCTTGTGCAAA 848

QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 849 AGGAGATGCTACTGTCAGTATTGTCATTCAGAACCAAGCAACATGTCAGGGGAAAGCTGG 908

QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 909 AGACAGCGCAGATATTATCATTTGCTCCGTTGGCAAGCAACATGTCAGGGGAAAGCTGG 968

QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 969 ATAAAGCCTGGTGCAGTCATTATTGACGTCGGAATCAACCCAGTAGAGGATCCAAATAGT 1028

QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1029 CCCGAGGTTACAACTGCTGGAGATGTTGTTATGAAGAGCCATAAGAATTCCTCTCT 1088

QY 261 LeuIleThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1089 GCTGTTACACCACTTCCTCGGAGGAGTGTGTCATGACCATAGCAATGCTTCTTCCAAAAT 1148

QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1149 ACATCTATTCTGCCAAGAGGCTGACCATTTTGAA 1184
```

RESULT 4

```
US-09-903-814-19
; Sequence 19, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 19
; LENGTH: 926
; TYPE: DNA
; ORGANISM: Glycine max
US-09-903-814-19
```

```
Alignment Scores:
Pred. No.: 3,35e-120 Length: 926
Score: 1030.00 Matches: 194
Percent Similarity: 89.63% Conservative: 22
```

```
Best Local Similarity: 80.50% Mismatches: 25
Query Match: 69.31% Indels: 0
DB: 9 Gaps: 0
US-10-723-061-18 (1-292) x US-09-903-814-19 (1-926)

QY 52 LysLysLysAlaCysGluAlaValGlyLeuSerTyrGluValAsnLeuProGluAsp 71
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 6 GAGAAGAAGGCTTGTGAATCTGTGGAATCAATCTTTGGAAGCAATCTGCTGAGAAT 65

QY 72 SerSerGluAspGluValLeuLysHisLeuAlaThrPheAsnSerAspProSerValHis 91
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 66 TCCACAGAAGAAGAGTTTGAATATATTGAGGCTACAAATGATGATGATGATGATGAT 125

QY 92 GlyLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnLeuLeuAsnAla 111
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 126 GGCATCTCTGTTGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 185

QY 112 ValSerLeuGluLysAspValAspGlyPheHisProLeuAsnLeuGlyArgLeuAlaMet 171
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 186 GTCCAGGATTGAGAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 245

QY 132 GlnGlyArgAspProPheValProCysThrProLysGlyCysMetGluLeuLeuHis 151
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 246 CGTGGAGAAGACCTCTGTTGTTTCTTGTACCAAGGAGATGATGATGATGATGAT 305

QY 152 ArgTyrGlyValGluLeuLysGlyLysArgAlaValValLeuGlyArgSerAsnLeuVal 171
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 306 AGATACAACTGTTCTATTAAGGAAGAGGCTGTTGATGATGATGATGATGATGAT 365

QY 172 GlyMetProAlaAlaLeuLeuGlnLeuLysAlaAsnAlaThrValSerLeuValHisSer 191
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 366 GGAATGCCAGTCTCTCTGTTTCAAGGGAAGATGCTACTGTCAGTATGTTCCATCT 425

QY 192 AsnThrLysLysProGluLeuThrArgGlnAlaAspLeuValLeuAlaValGly 211
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 426 AGAACCAAGTAAACCCGAGAGATCAATAGACAGGAGATGATGATGATGATGATGAT 485

QY 212 ValAlaAsnLeuValArgGlySerTyrPheLysProGlyValAlaLeuValGly 231
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 486 CAAGCAAAACATGTTGAGGGAAGCTGATAAAGCTGTTGTCAGTATGATGATGATGAT 545

QY 232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyValCys 251
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 546 ATCAACCCGGTAGAGATCAAAATAGCCCCCGAGGTTTCAAACTGGTGGAGATGTTGT 605

QY 252 TyrGluGluAlaSerLysLeuAlaGlyLeuLeuThrProValProGlyValGlyPro 271
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 606 TATGAAGAGCCATAAAGATTGCTCTGTCTTACCAAGGAGATGATGATGATGATGAT 665

QY 272 MetThrIleAlaMetLeuSerAsnThrLeuGluSerAlaLysArgLeuHisLysPhe 291
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 666 ATGACCATGACATGCTTCTCAAAATACACTCATCTCTGCAAGAGGAGGTCACAAATTT 725

QY 292 Lys 292
   ::::
Db 726 GAA 728

RESULT 5
US-10-723-061-19
; Sequence 19, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: B81179 USDIV1
; CURRENT APPLICATION NUMBER: US/10723, 061
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: US 60/092,869
```


Db 726 GAA 728

RESULT 6

US-10-425-114-33029

; Sequence 33029, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 33029

; LENGTH: 1066

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZMFLMO17055A08_FLI

US-10-425-114-33029

Alignment Scores:

Pred. No.: 1,79e-119 Length: 1066

Score: 1025.00 Matches: 192

Percent Similarity: 81.53% Conservative: 42

Best Local Similarity: 66.90% Mismatches: 53

Query Match: 68.98% Indels: 0

DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x US-10-425-114-33029 (1-1066)

Qy 1 AlalyslleleAspGlyLeuValAlalysGlnlleArgGluGluAlaValGlu 20

Db 25 GCGCAGATCATGTGGCAAGCCATCGCCGACGTCGCCGCGAGTCCGCCGCGAT 84

Qy 21 IleAlalysMetLysAspAlalleGlyValValProGlyLeuAlaValleLeuValGly 40

Db 85 GTGGCCGCGCTCTCGTGGCCACGAGCTCGTGGCGGGTGGCGGTGTCATGTTGGG 144

Qy 41 SerArgLysAspSerGlnThrThrValArgLysLysLysAlaCysGluAlaValGly 60

Db 145 AGCAGGAAGGACTCCGACAGCTACGTGAACATGAAGCCGCAAGCGTGGCGGAGTCCGC 204

Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80

Db 205 ATCTGCTCATCGAGTGTGACCTCCGCGAGGACATCTCCGAGCCGCTGTCGCCGAG 264

Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlylleLeuValGlnLeuProLeuPro 100

Db 265 GTTCATCGCTCAACGCTGACCCGCGTGCACGGATCCTTGTCCAGTTCCTCACTTCT 324

Qy 101 HisHisMetAsnAspGluAsnleLeuAsnAlaValSerlleGluLysAspValAspGly 120

Db 325 AAGCATATCAACAGAGAGAGATCTGACCGAGATTTCCATCGAGAAAGATGTGATGCG 384

Qy 121 PheHisProLeuAsnilleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140

Db 385 TTCCATCTCTCAACATTTGGCAAGCTTGCAATGAAGGACGAGACCACTGTCGTACCA 444

Qy 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLysGlyLys 160

Db 445 TGTACGCCAAGGGGTGATGAGTCTTGTCAAGGACGCGAGTCACTGTAAAGGTAAAG 504

Qy 161 ArgAlaValIleleArgSerAsnilleValGlyMetProAlaAlaLeuLeuGln 180

Db 505 CGGGCAGTGTGTTGGTGCAGCAACATCTGCGGGCTACCTGTATCCTCTCTCTCTG 564

Qy 181 LysAlaAsnAlaThrValSerlleValHisSerAsnThrLysLysProGluGluIleThr 200

Db 565 AAGCGCAGATCGCAGCGTATCTGTGTGCACTCGCGGACCCCTGATCTCTGAAGCATTTGA 624

Qy 201 ArgGlnAlaAspIleValleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220

Db 625 CGCGAAGCCGACATAGTCATCGCGCAGCTGGCGAGGTATGATGATCAAGGTGACTCG 684

Qy 221 IleLysProGlyAlaAlaIlelleAspValGlylleAsnProValAspAspProGluSer 240

Db 685 ATCAAGCCAGGTGCTCGGTCATCGATGTGCGGACGAACCTCGTCGACGACCTACCCCG 744

Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLyslleAlaGly 260

Db 745 AAGTCCGGGTACCGGCTCGTGGCGATGTGGATTTCGACGCGCGCGAGAGGTGCTGGG 804

Qy 261 LeuIleThrProValProGlyValGlyValProMetThrIleAlaMetLeuLeuSerAsn 280

Db 805 TACCTGACTCCGGTTCGCGAGGCGTTGCCCAATGACGGTGGCAATGCTGCTGAAGAAC 864

Qy 281 ThrLeuGluSerAlaLysArg 287

Db 865 ACGGTGGATGGGCAAGCGG 885

RESULT 7

US-10-425-114-23607

; Sequence 23607, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 23607

; LENGTH: 1415

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3597-067-D5_FLI

US-10-425-114-23607

Alignment Scores:

Pred. No.: 2,76e-119 Length: 1415

Score: 1025.00 Matches: 192

Percent Similarity: 81.53% Conservative: 42

Best Local Similarity: 66.90% Mismatches: 53

Query Match: 68.98% Indels: 0

DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x US-10-425-114-23607 (1-1415)

Qy 1 AlalyslleleAspGlyLysLeuValAlaLysGlnlleArgGluGluIleAlaValGlu 20

Db 406 GCGCAGATCATGTGGCAAGCCATCGCCGCGAGTCCGCCGCGAGTCCGCCGCGAT 465

Qy 21 IleAlalysMetLysAspAlalleGlyValValProGlyLeuAlaValleLeuValGly 40

Db 466 GTGGCCGCGCTCTCGTGGCCACGAGCTCGTGGCGGGCTGGCCGTGTCATCGTGGG 525

Qy 41 SerArgLysAspSerGlnThrThrValArgLysLysLysAlaCysGluAlaValGly 60

Db 526 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAAGCGTGGCGCGAGTCCGC 585

Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80

Db 586 ATCTGCTCCATCGAGTCGACCTCCCGAGGACATCTCCGAGACCGCGCTGTCGCCGAG 645
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleValGlnLeuProLeuPro 100
Db 646 GTTCATCGCTCAACGTGACCGCGCGGTGCACGGGATCCTTGTCCAGTTCACATTCCT 705
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 706 AAGCATATCAACGAGAGAGATACCTGAGCGAGATTTCCATCGAGAAAGATGTGGATGC 765
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db 766 TTCATCCTCTCAACATGTGCAAGCTTGCATGAAGAGGAGAGACCATGTTCGTACCA 825
Qy 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 826 TGTACGCCAAGGGGTGCATGGAGCTTGTCAAGAGCGGAGTCACTGTTAAAGTAAAG 885
Qy 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 886 CGGCGAGTGTGTGTCGACGACCAACATCGTCGGGCTACCTGTATCCCTGCTCTCTG 945
Qy 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 946 AAGCAGATCGACCGTATCTGTGTGCTCTCCGACCCCTGATCTGAAAGCATTTGA 1005
Qy 201 ArgGlnAlaAspIleValIleAlaValAlaGlyValAlaAsnLeuValArgGlySerTrp 220
Db 1006 CGCGAAGCGACATAGTCATCGCGGAGCTGGCGAGGCTATGATGATCAAGGTGACTGG 1065
Qy 221 IleIleProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 1066 ATCAAGCCAGTGTGCGGTTCATCGATGTCGGAGCAACTCGCTCAGCACCTACCCGG 1125
Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 1126 AAGTCGGGTACCGCTCGTCGCGATGTGGATTTCGACGCGCGAGCAAGGTGTGCTGG 1185
Qy 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 1186 TACCTGACTCCGGTTCGCGAGCGGTGGCCCAATCACGGTGGCAATGCTGCTGAAGAAC 1245
Qy 281 ThrLeuGluSerAlaLysArg 287
Db 1246 ACGGTGGATGGCGCAAGCGG 1266

RESULT 8

US-10-425-114-11296
; Sequence 11296, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 11296
; LENGTH: 1255
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701050070_FLI
; US-10-425-114-11296

Alignment Scores:

Pred. No.: 1.32e-118 Length: 1255

Score: 1019.00 Matches: 189
Percent Similarity: 81.66% Conservativity: 47
Best Local Similarity: 55.40% Mismatches: 53
Query Match: 58.57% Indels: 0
Gaps: 0
DB: 13

US-10-723-061-18 (1-292) x US-10-425-114-11296 (1-1255)

Qy 3 IleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGluIleAla 22
Db 312 GTGCTTGATGAAAGTTGATATCCATCGAATCAGATCAAAATAGCTGCTAGGTAAAGA 371
Qy 23 LysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGlySerArg 42
Db 372 CAATGAAGAAGCCCTAGGAAAGTTCTTGGATTAGCTGTAATCTTAGTGGGCCAAGA 431
Qy 43 LysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGlyIleLys 62
Db 432 AGGATCTCTAACTTACGTTTCGCAACAGATATATGCGCTGTGAAGAAGTTGAATCAAG 491
Qy 63 SerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAla 82
Db 492 TCTTTGGTGAATTAATCCACTGATGTGCGACTAACAGATGTTCAAAATGCCATCATG 551
Qy 83 ThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHis 102
Db 552 AGATTTAAACAAGATCCATTCATGCTGATGCTGTAGTGTGCTAGAAAAGATGTGGATG 611
Qy 103 MetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHis 122
Db 612 CTAGATGAGGAAAGTTTGGATGCTGTATGCTAGAAAAGATGTGGATGTTTCCAT 671
Qy 123 ProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThr 142
Db 672 CCCCTTAATATGGGAATCTTGCATTAAGAGAGGAGGCCACTGTTTACTCCTTGTACT 731
Qy 143 ProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysAla 162
Db 732 CCNAGGCTGTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 791
Qy 163 ValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLysAla 182
Db 792 GTAGTGTGCAACAAGTAATATTGGGTTCGAGCATCTTGTGTCAGAGACAC 851
Qy 183 AsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGln 202
Db 852 CATGCAACAGTCACTGTACATGCTTCCATCAACAAACCTGAAACAGATCACCTCAGAA 911
Qy 203 AlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrpIleLys 222
Db 912 GCTGATATTGTAGTTTCAGCTGCTGGAGTGCCTTAATTTGGTCCGTGGAACTGATAAAA 971
Qy 223 ProGlyAlaAlaIleIleAspValGlyLeuAsnProValAspAspProGluSerProArg 242
Db 972 CCTGGTGCAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1031
Qy 243 GlyTyrArgLeuValGlyAspValCysTyrGluAlaSerLysIleAlaGlyLeuIle 262
Db 1032 GGTATTCCCTCGCGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1091
Qy 263 ThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeu 282
Db 1092 ACTCCTGTGCTGGCGGGTGGCCCTATGACCCCTATGCTATGCTATGCTATGCTATG 1151
Qy 283 GluSerAlaLysArgIleHisLysPhe 291
Db 1152 GATTCTGCAAGCGCATGCTCAATTT 1178

RESULT 9

US-10-424-599-114288
; Sequence 114288, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

RESULT 10
US-09-903-814-15

Patent No. US20020102689A1
GENERAL INFORMATION:
APPLICANT: Falco, Carl
APPLICANT: Famodu, Layo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
APPLICANT: Thorpe, Cathy

```

; TITLE OF INVENTION:  LACTARYOZINASE METABOLISM ENZYMES
;
; FILE REFERENCE:  BB-1179-A
;
; CURRENT APPLICATION NUMBER:  US/09/903 814

```

; CURRENT FILING DATE: 2001-07-12

;; PRIOR APPLICATION NUMBER: 60/092,869

; PRIOR FILING DATE: 1998-07-15
 ;
 : NUMBER OF SEQ ID NOS. 22

```

; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97

```

; SEQ ID NO 15

; LENGTH: 1076

TYPE: DNA
ORGANISM: Zea mays

US-09-903-814-15

C
T
H
C
C
C
C
C

Alignment Scores:

Pred. No.:	2.51e-118	Length:	1076
score:	1016.00	Matches:	190

Score:	1016.00	Matches:	190
Percent Similarity:	81.18%	Conservative:	43

Percent Similarity: 81.15%
 Best Local Similarity: 66.20%
 Mismatches: 54

Query Match:	Indels:
68.37%	0

```
DB: 9 Gaps: 0
```

US-10-723-061-18 (1-292) x US-09-903-814-15 (1-1076)

(C) (C) (H) CT-ETG-COC-CO-CO-X (ZCC-T) GT-TGO-CZ/-AT-GC

Qy 1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluIleAlaValGlu 20

32 GCGACATCATCGATGCCAAGGCCATCGCGCGCGAGGTCCGCGCGGAT 91

DB 32 GUGCAGATCATCGATGGCAAGGCCATCGCGCGGACGTCGCGCGGAGGTCGCCCGGAT 51

Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

D_b 92 GTGGCCGGCTCTCGTCCGCCACGGACTCGTGCCGGGCTGGCCGTGTCATCGTGGG 151

41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60

[illegible]

D_b 152 AGCAGGAAGGACTCGCAGACGTACGTGAACATGAAGCGCAGGCGTGCGCGAGGTCGGC 211

[illegible]

QY I L E Y S S e r T y r G l u V a l A s n L e u P r o G l u A s p S e r G l u A s p G l u V a l L e u L y s H i s 80
61 I L E Y S S e r T y r G l u V a l A s n L e u P r o G l u A s p S e r G l u A s p G l u V a l L e u L y s H i s 80

Db
212 ATCTGCTCCATCGACGTCGACCTCCGGAGGACATCTCCGAGACCGGCTCGTCCGCGAG 271

[illegible]

Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100

272 GTTCATCGCCTCAACGCTGACCCCGCAGTGCA CGGATCCTTGTCAGGCTTCCACTTCCT 331

DD Z/2 911CA1C9CC1C8AC9C10ACCCCCCGCAGAGTGCACCGGAA1CC11GT1CCAC9C111CCAC111CC11

Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

332 AAGCATATCAACGAAGAGAAGATACTGAGCGAGATTTCCATCGAGAAAGATGTGGATGGC 391

121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
392 TTCCATCTCTCAACATTGCAAGCTTGCATGAAGGACAGAGCACTGTTCTGTACCA 451
141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLeuLysGlyLys 160
452 TGTACGCCAAGGGGTGCATGGAGCTCTTGTCAAGGAGCGGAGTCACTGTTAAAGGTAAG 511
161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
512 CGGCGAGTTGTGTGTCACGACATGCTCGGCTACCTGTATCCCTGCTCTCTCTG 571
181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
572 AAGCGACATGCGACCGGTATCTGTTGTGCACTCGGACCGCTGATCCCTCAAGACATTGTA 631
201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
632 CGGAGCTGACATGATGTCATCGGCGAGCTGGGACAGCTATGATGATCAAGGTGACTGG 691
221 IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
692 ATCAAGCCAGGTGCTGCGGTTCATCGATGTCGGACCACTCCATCGATGACCTACCCGG 751
241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
752 AAGTCCGGGTACGGCTGTCGGCGATGTGGATTTCGACGCGCGGACGCAAGGTGCTGGG 811
261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
812 TACTGATCCGGTTCCCGGAGCGTTGGCCCAATGACGCTGCGCAATGCTGCTGAAGAAC 871
281 ThrLeuGluSerAlaLysArg 287
872 ACGGTGATGGGCAAGCGG 892

RESULT 11

US-10-723-061-15
; Sequence 15, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
; APPLICANT: Faico, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 US/DivI
; CURRENT APPLICATION NUMBER: US/10723,061
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: US 09/903,814
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 1076
; TYPE: DNA
; ORGANISM: Zea mays
US-10-723-061-15

Alignment Scores:
Pred. No.: 2,51e-118 Length: 1076
Score: 1016.00 Matches: 190
Percent Similarity: 81.18% Conservative: 43
Best Local Similarity: 66.20% Mismatches: 54
Query Match: 68.37% Indels: 0
DB: 17 Gaps: 0

US-10-723-061-18 (1-292) x US-10-723-061-15 (1-1076)

1 AlalysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
32 CGCGAGATCATGATGCAAGGCATCGCGCCGACGTCGCGCGAGGTCGCGCGCGAT 91
21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
92 GTGGCCGCTCTCTCGTCCGCCACGAGCTGTCGCGGGGTGCGCGGTGTCATCGTGGG 151
41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
152 AGCAGGAGGACTCGACAGCTGACGTGAACATGAGCGCAAGCGGTCCGCCGAGTCCGC 211
61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
212 ATCTGCTCCATCGACGTCCGACCTCCGAGAGACATCTCCGAGACCGGCTCGTCCGCGAG 271
81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
272 GTTCATCGCTTCAACGCTGACCGCGAGTGCAGGGATCCTTGTCCAGCTTCCACTTCT 331
101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
332 AAGCATATCAACGAGAGAGATCTGAGCGAGATTTCATCGAGAAAGATGTGGATGGC 391
121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
392 TTCCATCTCTCAACATTGGCAAGCTTGCATGAAGGACGAGACCACTGTTTCGTACCA 451
141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLysGlyLys 160
452 TGTACGCCAAGGGTGCATGGAGCTTGTTCAGGAGCGGAGTCACTGTTAAAGGTAAG 511
161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
512 CGGCGAGTTGTGTGTCGACCAACATCGTCCGGCTACCTGTATCCCTGCTCTCTCTG 571
181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
572 AAGCGATGCGACCGCTGTTGTGTCATCGGACCGCTCGGACCGCTGATCTGAAAGCATGTA 631
201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
632 CGCGAAGCTGACATAGTCAATCGCGGACGCTGGGCGAGCTATGATGATCAAGGTGACTGG 691
221 IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
692 ATCAAGCCAGGTGCTCGGTCATCGATGTCGGGACGAACTCCATCGATGACCTACCCGG 751
241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
752 AAGTCCGGGTACGGCTCGTCCGCGATGTGGATTTCGACGCGCGGACGCAAGGTGCTGGG 811
261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
812 TACTGACTCCGGTTCGCGAGCGGTGGCCCAATGACGCTGCGCAATGCTGCTGAAGAAC 871
281 ThrLeuGluSerAlaLysArg 287
872 ACGGTGATGGGCAAGCGG 892

RESULT 12

US-10-437-963-74482
; Sequence 74482, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Fihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 74482
LENGTH: 1356
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_74664C.1
US-10-437-963-74482

Alignment Scores:
Pred. No.: 1,15e-117 Length: 1356
Score: 1012.00 Matches: 191
Percent Similarity: 80.84% Conservative: 41
Best Local Similarity: 56.55% Mismatches: 55
Query Match: 58.10% Indels: 0
DB: 17 Gaps: 0

US-10-723-061-18 (1-292) x US-10-437-963-74482 (1-1356)
Qy 1 AlaLysIleAlaAspGlyLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 99 GCGAGATCATCGACGGAAGGGGTGCGCCGACATCCCGCGAGTCCGCGCGAC 158
Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 159 GTCGCGCCCTCTCTCCGCGCCCAACCTCTGTCGCGCGGTGGGTGATGTCGTTGGG 218
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
Db 219 AGCAGGAAGGACTCGACAGTACGTGCGAGATGAAGCGCAAGCGTGGCGAGTCCGC 278
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerGluAspGluValLysHis 80
Db 279 ATCCGCTCGTCGAGTCTGCTGCGAGACATCTCCGAGCGGGCTGCTGCGCGAG 338
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 339 GTCCACCCGCTCAACGCGCGACCCAGCGCTCCATGCGATCTGTCAGTTCCTTCCCA 398
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 399 AAGCATATTATAGGAAGAAAGATATTGAACGAGATCAGCTTAGAAGAAAGATGTTGATGG 458
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 459 TTCATCTCTGAATATTGGCAAGCTTGCAATGAAGCAGACAGACCCACTGTTCTTCCA 518
Qy 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyHis 160
Db 519 TGCACGCGCAAGGGATGATGAGCTCTCTAACGCGAGCGAGATACCATCAATGGGAAG 578
Qy 161 ArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 579 CGAGCGGTCTGTTGGCGCGAGACATCTGCGGCTACTGTATCCCTGCTTCTCTG 638
Qy 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 639 AAGCGGGATGCGACCGTATCTTCTCACTCCCGGACCCCAACCCCTGAAAGTATTGTC 698
Qy 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db 699 CGTGAAGCAGACATTGTCTATGCGCGGCTGCCAGGTATGATGATCAAGAGACATGG 759
Qy 221 IleLysProGlyAlaAlaIleAlaIleAspValGlyIleAsnProValAspProGluSer 240
Db 759 ATCAAACCGCGCTGCTGTCATGCGACGCTTGGAGCGAACTCCATCAGTGACCCCAACGAG 818

Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 819 AAATCGGGGTACAGACTCTCGCGCATGTGGATTTCCGACAGGTGAGCAAGGTGCTGT 878
Qy 261 LeuIleThrProValProGlyValGlyValProMetThrIleAlaMetLeuLeuSerAsn 280
Db 879 CACCTGACTCCAGTCCAGTCCAGTGGCGTGGCCCATGACGCTGGCGATGTTGCTGAAGA 938
Qy 281 ThrLeuGluSerAlaLysArg 287
Db 939 ACGGTGGATGGAGCGAAACGT 959

RESULT 13
US-10-424-599-125693
Sequence 125693, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 125693
LENGTH: 1286
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_8450C.1
US-10-424-599-125693

Alignment Scores:
Pred. No.: 1.9e-117 Length: 1286
Score: 1010.00 Matches: 188
Percent Similarity: 81.88% Conservative: 47
Best Local Similarity: 65.51% Mismatches: 52
Query Match: 67.97% Indels: 0
DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x US-10-424-599-125693 (1-1286)
Qy 1 AlaLysIleAlaAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 64 GCCACCGTAAATCGACGCGCAAGCGTTCGCGCAACCATCCGATCTGAAATCGCGACGAG 123
Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 124 GTCCGCCAGCTCTCTCAAAAATACGCAAGGTTCCGGGATTAGCGTGTGTAGTAGGG 183
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
Db 184 AACAGAAAAGATTACAGAGCTATGTTGGAATGAAGAGAAAGCGATCCGCTGAATTTGGA 243
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 244 ATCAATCTCTCGATGTGGACCTTCCCGAACAGTCTCCAGCTGAATCAACAA 303
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 304 GTTCACGAGTTGAATGCTAACCTGACGTACATGTTATATGTTTCCGCTCCGCTTGCCT 363
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 364 AAGCATAAATGAAGAGGAGTTCGATGAAATCGACCTTGAAGAGGATGTAGATGT 423
Qy 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 424 TTTCATCTTTGAACATTGGCAAGCTTGAATGAAGGCGAGAGACCCCTGTTCTCTTCCC 483

; SEQ ID NO 23018
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_28138C.1
US-10-437-963-23018

Alignment Scores:
Pred. No.: 1,34e-114 Length: 1413
Score: 988.00 Matches: 186
Percent Similarity: 80.77% Conservative: 45
Best Local Similarity: 65.03% Mismatches: 55
Query Match: 66.49% Indels: 0
DB: 17 Gaps: 0

US-10-723-061-18 (1-292) x US-10-437-963-23018 (1-1413)

```
QY 1 AlaLysIleIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
324 GCGACCAATTATGATGGCAAGTCTGTCCGGAAGACATAAGGTTTCAGATCGCTGAAGAA 383
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 GTTCGTCAATGAAATGCAAGTGGGCGATGTGCTGGCCTAGCTGTGTATTGTTGGTCGC 443
QY 41 SerArgLysAspSerGlnThrTyxValArgAsnLysLysAlaCysGluAlaValGly 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
444 GATAGAGGACTCTGAGTCATACGTGCGATACAAATTAAGGCTGTGAGGAGTTTGA 503
QY 61 IleLysSerTyxGluValAsnLeuProLysAspSerSerGluAspGluValLeuLysHis 80
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
504 ATAAAGTCCTTGTGGCGGAGTTGCTGGGAACCTGACGGAAGACGTAGTGGTATTCT 563
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
564 GTGTCAAGATTCAATGAAGATCCATCTCTTCATGGCATTTCTTGCACTACCTCTACCA 623
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
624 CAGCATATGGATGAAGAAAGGATTCTGAGCGCTATTAGCCTAGAGAAGGATGTTGATGGT 683
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
684 TTCCATCCATTGAATGTTGTAATCTTGCCCTCAGAACCGGAAGCCCTTTATTGTGCC 743
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyxGlyValGluIleLysGlyLys 160
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
744 TGTGCTGCAAGGCTTGCTTGAGTTATTGCTCCAGTCTGGGATTGAACCTCATGGGAAA 803
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
804 CATGTCACCTGTGATTGGGAGAGCAAGTTGTGGTTGCCCACTCTTTACTTTTACAG 863
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
864 AGACACCATGCCACTGTTAGTATTATCCATGCTTTCCACAAATCCAGAGAGATTACC 923
QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
924 CGTCAATCTGATTGTGATCTCAGCTGCGAGTGGCAATCTTGTAAAGAGGAAGCTGG 983
QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
984 TTGAAGAAAGTGCAGTTGTGATGATGTTGGAACAACCAATGAGGATCCCAACCACT 1043
QY 241 ProArgGlyTyxArgLeuValGlyAspValCysTyxGluGluAlaSerLysIleAlaGly 260
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1044 GATTACGGCTATCGATTAACTGGAGATGTCTGCTTCGAAGAAGCTGTGAAGTGGCTCT 1103
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1104 GCTATAACTCCAGTCTCTGGTGGTGGACCGGTGACCATTTGGATGCTTCTTGGCAAT 1163
```

QY 281 ThrLeuGluSerAlaLys 286
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1164 ACACTTGACTCAGCGAAA 1181

Search completed: July 24, 2004, 11:12:47
Job time : 587 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 24, 2004, 06:18:40 ; Search time 59 Seconds
(without alignments)
476.067 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIIDGKLVAKQIREIEAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	83.3	360	T01226	probable methylene
2	1034	69.6	294	T50664	methylenetetrahydr
3	1017	68.4	350	G84807	methylenetetrahydr
4	1009	67.9	310	T01224	probable methylene
5	763.5	51.4	284	A21607	methylenetetrahydr
6	755.5	50.8	284	AH1244	methylenetetrahydr
7	749.5	50.4	937	T40723	C-1-tetrahydrofola
8	741	49.9	319	A23573	methylenetetrahydr
9	739.5	49.8	946	A23550	methylenetetrahydr
10	735	49.5	935	A31903	methylenetetrahydr
11	730	49.1	344	C87400	Fold bifunctional
12	724	48.7	285	D50395	hypothetical prote
13	718	48.3	935	A35367	methylenetetrahydr
14	716	48.2	279	H83997	methylenetetrahydr
15	715	48.1	306	A37963	methylenetetrahydr
16	713.5	48.0	975	A28174	methylenetetrahydr
17	712	47.9	933	S53523	methylenetetrahydr
18	707.5	47.6	294	E86731	hypothetical prote
19	704	47.4	284	F82031	methylenetetrahydr
20	689	46.4	282	G81358	methylenetetrahydr
21	683.5	46.0	283	E69626	methylenetetrahydr
22	681.5	45.9	282	A64081	methylenetetrahydr
23	679	45.7	288	J50662	methylenetetrahydr
24	677	45.6	284	H83421	5,10-methylene-tet
25	676	45.5	285	S71924	methylenetetrahydr
26	674	45.4	288	G50702	5,10-methylene-tet
27	674	45.4	288	B55533	5,10-methylene-tet
28	673	45.3	288	I36633	methylenetetrahydr
29	673	45.3	288	AF0569	hypothetical prote

30	668	45.0	291	2	F70463	methylenetetrahydr
31	668	45.0	299	2	AD1833	methylenetetrahydr
32	667	44.9	299	2	AH2648	hypothetical prote
33	667	44.9	319	2	G97430	hypothetical prote
34	665.5	44.8	286	2	E89875	Fold bifunctional
35	662	44.5	969	2	T40147	probable tetrahydr
36	661	44.5	311	2	H82136	methylenetetrahydr
37	660	44.4	288	2	A80344	methylenetetrahydr
38	652.5	43.9	357	1	S32562	methylenetetrahydr
39	648	43.6	297	2	A84296	probable methylene
40	632	42.5	315	2	A71288	methylenetetrahydr
41	629.5	42.4	350	1	A33267	methylenetetrahydr
42	627.5	42.2	306	2	T51705	methylenetetrahydr
43	626	42.1	295	2	C82558	bifunctional methy
44	623.5	42.0	344	1	DEHUMT	methylenetetrahydr
45	621	41.8	294	2	S76287	5,10-methylene-tet

ALIGNMENTS

RESULT 1

T01226

probable methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) F6N23.26 - Arabidops

N;Alternate names: protein F6N23.26

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002

C;Accession: T01226

R;Geisel, C. submitted to the EMBL Data Library, April 1998

A;Description: The sequence of A. thaliana F6N23.

A;Reference number: Z14281

A;Accession: T01226

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-360 <GEI>

A;Cross-references: EMBL:AF058919; NID:G3047100; PID:G3047116; GSPDB:GN00063; ATSP:F6N23.

A;Experimental source: cultivar Columbia

C;Genetics:

A;Gene: ATSP:F6N23.26

A;Map position: 5

A;Introns: 61/1; 169/3; 248/3; 304/3

C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

C;Keywords: NADP; oxidoreductase

F;71-354/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <WTFD>

Query Match 83.3%; Score 1238; DB 2; Length 360;

Best Local Similarity 79.1%; Pred. No. 1.1e-90;

Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAKQIREIEAVEIAKMDAICGVPLAVILVGSRKDSQTYVRNKKACEAVG 60

Db 69 AIVIDGKAVAKIRDEITIEVRMKESIGVIPGLAVILVGSRKDSATYVRNKKACDSVG 128

QY 61 IKSYENLPLDSDSEVLKHIATFNSDPVHGILVQLPLPHMNDENILNAVSIKDVVG 120

Db 129 IKSFVRLEADSEEVLSKVSFGFNDPSVHGILVQLPLPSHWDENILNAVSIKDVVG 188

QY 121 FHLPLNIGRAMQGRDFFVPCPKGMELLHYGVEIKGKRAVIGRSNIVGMPAALLIQ 180

Db 189 FHLPLNIGRAMGREPLFVPCPKGCIELHRYNIEIKGKRAVIGRSNIVGMPAALLIQ 248

QY 181 KANATSVIHSNTKPEETITROADIVIAVGVAIVRGSMWIKPGAIDVGINPVDDPS 240

Db 249 REDATSVIHSRTKPEETITREADIIISAVGPNVRGSIKFGAVLIDVGINPVDDPSA 308

QY 241 PRGYRLVGVDCVEEASKIAGLITPVPGVGPMTIAMLLSNTLESAKRIHKFK 292

Db 309 ARGVRLVGVDCVEEASKVASAITPVPGVGPMTIAMLLSNTLTSKRIHNFQ 360

RESULT 2

T50664

```

123 SVLAELPEDCTEGQIIIVLAKFNEDTSIHGILLVOLPLPQHLNESKIILMWVLEKVDVGFH 182
123 PLNIGRLAMQGRDPFFVFCPTKGCWELLHRYGVBEIKGRAVWIGRSNIVGMPAALLQKA 182
183 PLNVGNLAMRGREFLFSCTPTKGCVELLITRTGVEIAKNAVWIGRSNIVGLPWSLLILQRH 242
183 NATVSIVHSNTKPEEITROADIIVAAVGVANLVGRGSKIWIKPGAALIDVGINPVDDPSPR 242
243 DATVSTVHAFTKDEHITRKADIIVAAAGIPNLVRGSLKPGAVVIDVGTTPVEDSCEF 302
243 GYRLVGVDCVVEEASKIAGLITPVPGGVGPMTIAMLLSNTLESAKRI 288
303 GYRLVGVDCVVEEALGVASAITPVPGGVGPMTIITMLLCNTLEAAKRI 348

RESULT 4
T01224
probable methylentetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) F6N23.28 - Arabidops
N:Alternate names: protein F6N23.28
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 03-Jun-2002
C:Accession: T01224
R:Geisel, C.
submitted to the EMBL Data Library, April 1998
A:Description: The sequence of A. thaliana F6N23.
A:Reference number: Z14281
A:Accession: T01224
A:A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-310 <GEI>
A:Cross-references: EMBL:AF058919; NID:g3047100; PID:g3047121; GSPDB:GNC0063; ATSP:F6N23.
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:F6N23.28
A:Map position: 5
A:Introns: 87/1; 119/3; 198/3; 254/3
C:Superfamily: methylentetrahydrofolate dehydrogenase (NAD+); methylentetrahydrofolate
C:Keywords: NADP; oxidoreductase
F:23-304/Domain: methylentetrahydrofolate dehydrogenase (NAD+) homology <MTFD>

```

```

Query match: 0.78; Score 100%; E=0.00; Length 310;
Best Local Similarity 66.4%; Pred. No. 1.3e-72; Indels 36; Gaps 1;
Matches 194; Conservative 29; Mismatches 29;
1 AKTIIDGKLVAKQRIREEIAVIAKMKAIDGVVPGVLAVILGSRDQSQTIVANKKKCAEAVG 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
55 AIVIDGKAEEKIRDDIKIEVRMKESIGVVPA----- 87

        61 IKSVEYNLPEDSSEDEVLKHIAFNSPDSYHGHTILVOLPLPHMNDENILNAVSIKDVG 120
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
88 -----EDSSSEEVKTVSGFNDPDSYHGVLVQLPLPSHMDSEQNILNAVSIKDVG 138

        121 FHLNIGRLAMQGRDPFFVCTPKGCMELIHRHYGEIKGRAVVGISNVIGMPAALLQ 180
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
139 FHLNIGRLAMRGREPLVFCTPKGCIELLHRYNIETFGKRAVVGISNVIGMPAALLQ 198

        181 KANATVISVSNTSKKPZEITRQADIVTAAAGVANLRVGSWIKPCAALIIDVGINPVDDPES 240
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
199 KEDATVSIHSRTWNPELITRQADILLISAVGKNWRGSWIKPCAVLIDVGIKPEDPSA 258

        241 PRGYRLVDVCYEBASKIAGLIITVPVGCGFMPTIAMLLSNLTLSAKRIHKFK 292
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
259 AGGERLVGDICYEBASKIASAITVPVGDMGTMIAMLLSNLTLSAKRIHNFK 310
           : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
RD1607
methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase homol
C:Species: Listeria innocua
C:CDate: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1607

```

D.; Jones, L.M.; Karst, U.
Science 294: 849-852, 2001
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A./Title: Comparative genomics of *Listeria* species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AD1607
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-284 <GLA>
A/Cross-references: GB:AL592022; PIDN:CAC96628.1; PID:g16413870; GSPDB:GN00178
A/Experimental source: strain Clip11262
C/Genetics:
A/Gene: fold
C/Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 51.4%; Score 763.5; DB 2; Length 284;
Best Local Similarity 52.2%; Pred. No. 3.7e-53;
Matches 151; Conservative 59; Mismatches 70; Indels 9; Gaps 4;
Qy 2 KIIDGKLVAKOIREBIAEIAEAKMDAIGVWPGGLAVILVGRKDSQTYVRNKKKACAEAVGI 61
Db 3 EIIDGKLVAKOIREBIAEIAEAKMDAIGVWPGGLAVILVGRKDSQTYVRNKKKACAEAGM 61
Qy 62 KSYEVNLPEDSSDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVDF 121
Db 62 KSVLIELPETVTEELKLVVEELNEDTTHIGILVOLPLPHKHSIEKVIDAISFDKVDGF 121
Qy 122 HPLNIGRLAMQGRDPFFVPCPKGCMELLHRYGVEIKGKRAVIGRSNIVGMPAAALLQK 181
Db 122 HPVNVGNLFI-GKDS-FVPCPTAGIILIKSTGTQIEGKRAVIGRSNIVGKPAQALLN 179
Qy 182 ANATSVIHSNTKPEEITROADIVIAAVGVANLVGRSWIKPGAAIIVDGINPDDPESP 241
Db 180 ENATVTIAHSRTKOLPQVAKEDILVATGLAKFKVKYIKPGAIVIDVGM-----DRD 233
Qy 242 RGYRLVGDVCEYEASAKIAGLITPVGPGVPMTIAMLLSNTLESASAKRIHK 290
Db 234 ENNKLCGVDVDFDVKQAGFITPVGPGVPMTIAMLLANTLKAARKRIWK 282

RESULT 6
AH1244
methylenetetrahydrofolate dehydrogenase and methylenetetrahydrofolate cyclohydrolyase hom
C/Species: *Listeria monocytogenes*
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C/Accession: AH1244
R/Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294: 849-852, 2001
A/Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A./Title: Comparative genomics of *Listeria* species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AH1244
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-284 <GLA>
A/Cross-references: GB:NC_003210; PIDN:CAC99438.1; PID:g16410776; GSPDB:GN00177
A/Experimental source: strain EGD-e
C/Genetics:
A/Gene: fold
C/Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 50.8%; Score 755.5; DB 2; Length 284;
Best Local Similarity 51.6%; Pred. No. 1.6e-52;
Matches 149; Conservative 60; Mismatches 71; Indels 9; Gaps 4;
Qy 2 KIIDGKLVAKOIREBIAEIAEAKMDAIGVWPGGLAVILVGRKDSQTYVRNKKKACAEAVGI 61
Db 3 EIIDGKLVAKOIREBIAEIAEAKMDAIGVWPGGLAVILVGRKDSQTYVRNKKKACAEAGM 61

Qy 62 KSYEVNLPEDSSDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVDF 121
Db 62 KSVLIELPETVTEELKLVVEELNEDTTHIGILVOLPLPHKHSIEKVIDAISFDKVDGF 121
Qy 122 HPLNIGRLAMQGRDPFFVPCPKGCMELLHRYGVEIKGKRAVIGRSNIVGMPAAALLQK 181
Db 122 HPVNVGNLFI-GKDS-FVPCPTAGIILIKSTGTQIEGKRAVIGRSNIVGKPAQALLN 179
Qy 182 ANATSVIHSNTKPEEITROADIVIAAVGVANLVGRSWIKPGAAIIVDGINPDDPESP 241
Db 180 ENATVTIAHSRTKOLPQVAKEDILVATGLAKFKVKYIKPGAIVIDVGM-----DRD 233
Qy 242 RGYRLVGDVCEYEASAKIAGLITPVGPGVPMTIAMLLSNTLESASAKRIHK 290
Db 234 ENNKLCGVDVDFDVKQAGFITPVGPGVPMTIAMLLANTLKAARKRIWK 282

RESULT 7
T40723
c-1-tetrahydrofolate synthase - fission yeast (Schizosaccharomyces pombe)
C/Species: *Schizosaccharomyces pombe*
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2000
C/Accession: T40723
R/Lyme, M.; Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Saund
submitted to the EMBL Data Library, July 1999
A/Reference number: Z21946
A/Accession: T40723
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-937 <LYN>
A/Cross-references: EMBL:AL096796; PIDN:CAB46709.1; GSPDB:GN00067; SPDB:SPBC839.16
A/Experimental source: strain 972h; cosmid c839
C/Genetics:
A/Gene: SPDB:SPBC839.16
A/Map position: 2
A/Introns: 12/2; 60/3
C/Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology <FTL>
F:312-937/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 50.4%; Score 749.5; DB 2; Length 937;
Best Local Similarity 50.0%; Pred. No. 2.3e-51;
Matches 147; Conservative 50; Mismatches 90; Indels 7; Gaps 2;

Qy 1 AKTIIDGKLVAKOIREBIAEIAEIAKMDAIGVWPGGLAVILVGRKDSQTYVRNKKKAC 56
Db 2 ALLEGTSLARKVREELRQISIK---SVDPYFVNSLKIIVGGEDSNVVRKTRAA 58
Qy 57 EAVGKSYEVNLPEDSSDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIK 116
Db 59 NEAGISCEHVNFPEITVYDILLAIKGFNEPTVHGIIVOLPLPAHINEQIITEVAPEK 118
Qy 117 DVDGPHPLNIGRLAMQGRDPFFVPCPKGCMELLHRYGVEIKGKRAVIGRSNIVGMPAA 176
Db 119 DVDGFCETNLGLKLTKEGGPLFTACTPKGIMCLIKHGINVOGKRAVIGRSNIVGRMS 178
Qy 177 LILOKANATVSIVHSNTKPEEITROADIVIAAVGVANLVGRSWIKPGAAIIVDGINPVD 236
Db 179 ILLEKANATVTLCHSKTESIADIVRTADIVAAIGIPHFVKADLKKGVVAIDVGINSIP 238
Qy 237 DPSPRGYRLVGDVCEYEASAKIAGLITPVGPGVPMTIAMLLSNTLESASAKRIHK 290
Db 239 DATKKSGLTGDIDFENAKEVASAIPVPGSGVPMTIAMLLQNVSASVAVRFRK 292

RESULT 8
AE3573
methylenetetrahydrofolate cyclohydrolyase (EC 3.5.4.9) [imported] - *Brucella melitensis* (st
C/Species: *Brucella melitensis*
C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C/Accession: AE3573
R/DelVecchio, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova, I
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen *Bruceella melitensis*
A;Reference number: AD3252; PMID:11756688
A;Accession: AE3573
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <CUR>
A;Cross-references: GB:AE008918; PIDN:AAL53752.1; PID:g17984679; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI10510
A;Map position: 11
C;Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C;Keywords: hydrolase

Query Match 49.9%; Score 741; DB 2; Length 319;
Best Local Similarity 51.6%; Pred. No. 2.6e-51;
Matches 149; Conservative 49; Mismatches 69; Indels 2; Gaps 2;

QY 1 AKIIDGLVAKQIREIAVEIAKMDAIGVPGGLAVILVGSRKDSQTYRNKKKACEAVG 60
DB 22 AQIDGKKAEDVSVSTKTEKLVAATGVPGIAVIVGEDPASQVYVASKSRKAKCG 81
QY 61 IKSEYVNLPSDESEDEVLKHIAIFNSDPSVHGLVQLPLPHMNDENILNAVSIKDVVG 120
DB 82 FHSVQHDLPETASEQELLNLLEGLNNDPAIHGLVQLPLPHGIDSGRVIQTIAPEKDVVG 141
QY 121 FHPINTGRLAQGRDPFFVPCPKGCWELLHR-YGVEIKGKAIVIGRSNIVGMPAALL 179
DB 142 FHPINVGKLTGTVETAFVPCCTPAGAMIMIERVHGRLDGLNAVIGRSNIVGKPMFNL 201
QY 180 QKANATVSIHNSNTKPEETITQADIVIAAVGVANLVRSWKPGAAIIVDGINVDPE 239
DB 202 LAANATVVAHSTKLPALARNADILVAAGRPQVKGWVKGATVIVDGINRIPAPE 261
QY 240 SPRG-YRLVGDVVCYERASKIAGLITPVPGVGPMTIAMLLSNTLSAKR 287
DB 262 RGEKTRLVGDVDFAEAKVAGAITVPVGGVGPMTIAMLMANTLTAACR 310

RESULT 9

A29550
methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) - yeast (*Saccharomyces cerevisiae*)
N;Alternate names: C1-tetrahydrofolate synthase; protein G7733; protein YGR204W
N;Contains: formate-tetrahydrofolate ligase (EC 6.3.4.3); methylenetetrahydrofolate cyclase
C;Species: *Saccharomyces cerevisiae*
C;Date: 21-May-1988 #sequence_revision 05-Apr-1995 #text_change 19-Jul-2002
C;Accession: A29550; S33927; S64526; S63853
R;Staben, C.; Rabinowitz, J.C.
J. Biol. Chem. 261, 4629-4637, 1986
A;Title: Nucleotide sequence of the *Saccharomyces cerevisiae* ADE3 gene encoding C1-tetra
A;Reference number: A92580; MUID:86168166; PMID:3514599
A;Accession: A29550
A;Molecule type: DNA
A;Residues: 1-946 <STA>
A;Cross-references: EMBL:ML2878; NID:g171004; PIDN:AAA6316.1; PID:g171005
R;Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the EMBL Data Library, April 1995
A;Description: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII
es, of the human.
A;Reference number: S53922
A;Accession: S53927
A;Molecule type: DNA
A;Residues: 1-946 <GUE>
A;Cross-references: EMBL:Z49133; NID:g790489; PIDN:CAA88997.1; PID:g790495
A;Experimental source: strain S228C
R;Guerrero, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64517
A;Accession: S64526
A;Molecule type: DNA
A;Residues: 1-946 <GUE>
A;Cross-references: EMBL:272989; NID:g1323364; PIDN:CAA97231.1; PID:g1323365; GSPDB:GN00
A;Experimental source: strain S288C

R;Guerrero, P.; Barreiros, T.; Soares, H.; Cyrne, L.; Maia e Silva, A.; Rodrigues-Pousa
Yeast 12, 273-280, 1996
A;Title: Sequencing of a 17.6 kb segment on the right arm of yeast chromosome VII reveals
terial electron-transferring flavoproteins (beta-chain) and of the *Escherichia coli* phos
A;Reference number: S63848; MUID:97060019; PMID:8904340
A;Accession: S63853
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-946 <GUP>
A;Cross-references: EMBL:Z49133; NID:g790489; PIDN:CAA88997.1; PID:g790495
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C;Genetics:
A;Gene: SGD:ADE3; MIPS:YGR204W
A;Cross-references: SGD:S0003436; MIPS:YGR204W
A;Map position: 7R
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; me
C;Keywords: hydrolase; ligase; multifunctional enzyme; NADP; one-carbon metabolism; oxid
F;5-297/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
F;322-946/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 49.8%; Score 739.5; DB 1; Length 946;
Best Local Similarity 50.5%; Pred. No. 1.4e-50;
Matches 149; Conservative 45; Mismatches 92; Indels 9; Gaps 2;

QY 2 KIDGKLVAQIREIAVEIAKMDAI-GVPGGLAVILVGSRKDSQTYRNKKKACEAVG 60
DB 4 QVLDGKCAQOFSRNTANEIKS:QGHVPGFAPNLAIQVNRPSATYVRMKKAAEAG 63
QY 61 IKSEYVNLPSDESEDEVLKHIAIFNSDPSVHGLVQLPLPHMNDENILNAVSIKDVVG 120
DB 64 IVANFIHLDESATEFEVLRVVDQNLNEDPHTHGIIIVQLPLPAHLDEDRITSRLAEKDVVG 123
QY 121 FHPINTGRLAQGRDPFFVPCPKGCWELLHRVYVIGKRAVIVIGRSNIVGMPAALL 180
DB 124 FGPTNIGELNKKGHFPFLCTPKGIIEHLKANVTIEGSRSVVIGRSDIVGSPVAEL 183
QY 181 KANATVSIHNSNTKPEETITQADIVIAAVGVANLVRSWKPG-----GAAIIVDGI 232
DB 184 SLNSTVTITHSKTRDIASVLDADIVVAIQGFVKGFEKPRDGTSSDKKTVIVDGT 243
QY 233 NPVDPSPGYRLVGDVVCYERASKIAGLITPVPGVGPMTIAMLLSNTLSAKR 287
DB 244 NYVADPSKSGFKCVGDVEFNEAIKYVHLITPVPGVGPMTIAMLLSNTLSAKR 298

RESULT 10

A31903
methylenetetrahydrofolate dehydrogenase (NADP) (EC 1.5.1.5) - human
N;Contains: formate-tetrahydrofolate ligase (EC 6.3.4.3); methylenetetrahydrofolate dehy
C;Species: *Homo sapiens* (man)
C;Date: 21-May-1990 #sequence_revision 05-Apr-1995 #text_change 19-Jul-2002
C;Accession: A31903
R;Hum, D.W.; Bell, A.W.; Rozen, R.; Mackenzie, R.E.
J. Biol. Chem. 263, 15946-15950, 1988
A;Title: Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding
ase.
A;Reference number: A31903; MUID:89034046; PMID:3053686
A;Accession: A31903
A;Molecule type: mRNA
A;Residues: 1-935 <HUM>
A;Cross-references: GB:J04031; NID:g187464; PIDN:AAA59574.1; PID:g307178
C;Genetics:
A;Gene: GDB:MTFHD1; MTHFD; MTHFC
A;Cross-references: GDB:120704; OMIM:172460
A;Map position: 14q24-14q24
C;Superfamily: C1-tetrahydrofolate synthase; formate-tetrahydrofolate ligase homology; me
C;Keywords: hydrolase; ligase; multifunctional enzyme; NADP; one-carbon metabolism; oxid
F;6-292/Domain: methylenetetrahydrofolate dehydrogenase (NAD+) homology <MTFD>
F;318-935/Domain: formate-tetrahydrofolate ligase homology <FTL>

Query Match 49.5%; Score 735; DB 1; Length 935;
Best Local Similarity 49.0%; Pred. No. 3.2e-50;
Matches 145; Conservative 56; Mismatches 91; Indels 4; Gaps 3;

QY 1 AKIIDGKLVAKQIREBIAVEIAKMWDAL-GVVPGLAVILVGRKKDSQTYYRNKKACEAV 59
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 4 AEILINGKETSAQIRADLRKNQVTQLKEQVGFPTPRLAILQVGNRRDSDNLINVKLKAAEBI 63
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 60 GIKSYEVNLPEDSSDEVLLKHATENSDFSVHGILVQLPL--PHNMNDENILNAVSIKD 117
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 64 GKIKATHIKUPRTTISEVWMKIITSLNEDSTVHGLFVLQPLDSENSINTSEVINAIPEKD 123
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 118 VDGFLPNTIGRLAMQGRDPFFVPCTPKGCMEILLHRYGVFIKGRRAVVGSRNVGMPPAAL 177
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 124 VDGLTSINAGRLARGDLNDCFIPCTPKGCLELTKETGTGPVIAGRHAIVVGRSKIVGAPMHD 183
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 178 LLOKANATVISVHSNTKKPBETITQADIIVAAGVANVRGSMWKPGKAALIDVGINPVDD 237
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 184 LLWMNNATVTCHSKTAHJDEEYNKGDLVVAGQPEMWKGEWIKPGAVIDCGINYPVD 243
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 238 PESFRGVRLLVGDVCYEASKIAGLIITFPVGGVGPMTIAMLLSNTLESAXR-IHKFX 292
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 244 DKXPNGRKVGDVAYDEAKERASFITFPVGGVGPMTVAMLQMSTVESAKEFLEKFX 299
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 11
C87400
Fold bifunctional protein [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 27-Nov-2001
C;Accession: C87400
R;Kiernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.B.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
N., J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, G.W.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: C87400
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-344 <STO>
A;Cross-references: GB:AE005673; NID:g13422543; PIDN:AAK23199.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1217
C;Superfamily: methylentetrahydrofolate dehydrogenase (NAD+); methylentetrahydrofolate

Query Match 49.1%; Score 730; DB 2; Length 344;
Best Local Similarity 51.7%; Pred. No. 2.2e-50;
Matches 151; Conservative 47; Mismatches 88; Indels 6; Gaps 3;

QY 1 AKIIDGKLVAKQIREBIAVEIAKMWDALGVVPGLAVILVGRKKDSQTYVNRKKACEAVG 60
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 51 AKLDGKFAADLRARAKIAEEVAALKABEGVTPGLVALVLVEDPASQYVVRNKGEQTTAAG 110
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 61 IKSYEVNLPEDSSDEVLLKHATNSDPSVHGILVQLPLPHNMNDENILNAVSIKDVGD 120
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 111 MYSETHLRPETTQDELLAVVAKLNADPKINGVLVQFPVPHISQMDWAALSFDKDVGD 170
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 121 FHPNIGRLAMQGRDPFFVPCTPKGCMEILLHRYGVFIKGRRAVVGSRNVGMPPAALLIQ 180
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 171 LTVTNAGRLA--SGLPALAPCTPGCMWLIRDATGDLKGTATVVGSRNLMGKPMQMLL 228
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 181 KANATVISVHSNTKKPBETITQADIIVAAGVANVRGSMWKPGKAALIDVGIN--PYDDP 238
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 229 AADCTVTIAHSRSQDLPSIVQAQDIIVAAGRAEMWKADWVKPAVIDVGITRFPAARDP 288
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 239 ESPRG--YRLVGDVCYEASKIAGLIITFPVGGVGPMTIAMLLSNTLESAKRI 288
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 289 EAAAAGKTRLVGDVAFDEVREVAAGITFPVGGVGPMTIACLLANTLTAAKRL 340
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 12
D95095
hypothetical protein SP0825 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence revision 03-Aug-2001 #text change 24-Aug-2001

Db 64 GIKATHIKLPRTSTSESVLKVISLNEADATVHGFIQVLPDSENSINTEAVINAIAPKD 123
Qy 118 VDGPHPLNIGRLAMQGRDPFFVPCPTPKGCVMLLHRYGVEIKGRVAVIGSRNIVGMPPAL 177
Db 124 VDLGTSINAGKLARDGLKDCFICTPKGCLLEIKETGVQVAGRAHVVGSRKIVGAPMD 183
Qy 178 LLOKANATVSVHNSNTKKPEITRQADIVIAAVGVANLVRSWIKPGAAIIVGINPVD 237
Db 184 LLLNNATVTTCSTKADLDKENVKGDILVVATQGPWVGKGEWIKPGAVVIDGGINVVD 243
Qy 238 PESRPGVRLVGDVCEYEASKIAGLITPVGCGVPMTIAMLLSNTLESAGR-IHKFK 292
Db 244 DTRNGRKKVGDVAYDEAKESAFITPVGCGVPMTIAMLLSNTLESAGR-IHKFK 299

RESULT 14
H83997
methylenetetrahydrofolate dehydrogenase fold [imported] - Bacillus halodurans (strain C-
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: H83997
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A/Reference number: A83650; MUID:20512582; PMID:11058132
A/Accession: H83997
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-279 <STO>
A/Cross-references: GB:AP001516; GB:BA000004; NID:G10175192; PIDN:BA06503.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: fold
C/Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate

Query Match 48.2%; Score 716; DB 2; Length 279;
Best Local Similarity 50.7%; Pred. No. 2.1e-49;
Matches 145; Conservative 52; Mismatches 79; Indels 10; Gaps 4;

Qy 1 AKIIDGKLVAKOIREIAVEIAKMDAIGVVGGLAVILVSGSKDSQTVYRNKKKACEAVG 60
Db 3 AVLLDGKLAARKDEKSEKSEVTKER-GITPLGAVILVGNPDASTVYVRSKQACEQIG 61
Qy 61 IKSEYVNLPESSSEDEVILKHIAIFNSDPSVHGILVQLPLPHMNDENILNAVSIKEDVDG 120
Db 62 IYVLKELPASTSEELLTEIDRLNNDPTIHGILVQLPLPEQISEQAVIERISPAKVDG 121
Qy 121 FHLNIGRLAMQGRDPFFVPCPTPKGCVMLLHRYGVEIKGRVAVIGSRNIVGMPPAL 180
Db 122 FHPISVGRM-MIGEDT-FLFCTPFGVLVMOEAVIEAGKVVVVGSRNIVGKPGVQLML 179
Qy 181 KANATVSVHNSNTKKPEITRQADIVIAAVGVANLVRSWIKPGAAIIVGINPVD 240
Db 180 NEHATVYCHSRKNTLTKADILVAVGARFIDASHVKEGAVIDVGINRVDG--- 236
Qy 241 PGRYLVGVCEYEASKIAGLITPVGCGVPMTIAMLLSNTLESAGR 286
Db 237 ----KLCDGVDFESVREVAHSLTPVPGGVPMTITMLLANTITQAAK 278

RESULT 15
A97963
methylenetetrahydrofolate cyclohydrolase (EC 3.5.4.9) fold [imported] - Streptococcus pne
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C/Accession: A97963
R/Hoskins, J.A.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234

A/Accession: A97963
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-306 <KUR>
A/Cross-references: GB:AE007317; PIDN:AAK99533.1; PID:G15458321; GSPDB:GN00174
C/Genetics:
A/Gene: fold
C/Superfamily: methylenetetrahydrofolate dehydrogenase (NAD+); methylenetetrahydrofolate
C/Keywords: hydrolase

Query Match 48.1%; Score 715; DB 2; Length 306;
Best Local Similarity 49.7%; Pred. No. 2.9e-49;
Matches 142; Conservative 54; Mismatches 82; Indels 8; Gaps 2;

Qy 2 KIIDGKLVAKOIREIAVEIAKMDAIGVVGGLAVILVSGSKDSQTVYRNKKKACEAVG 61
Db 24 QIIDGKALAKLQQLAEKTAKLEETALVPLGLVILVGNPASQVYRNKERSALAGS 83
Qy 62 KSEYVNLPESSSEDEVILKHIAIFNSDPSVHGILVQLPLPHMNDENILNAVSIKEDVDG 121
Db 84 RSEYVVRVPEITITQBEELDLIAKYNQDPAWHGILVQLPLPKHIDEEAVLLAIDPEKVDG 143
Qy 122 HPLNIGRLAMQGRDPFFVPCPTPKGCVMLLHRYGVEIKGRVAVIGSRNIVGMPPAL 181
Db 144 HPLNIGRL--WSGHVWIPSTPAGIMEMFHEYGIDLEGKNAVIGSRNIVGKPMQAQLLA 201
Qy 182 ANATVSVHNSNTKKPEITRQADIVIAAVGVANLVRSWIKPGAAIIVGINPVD 241
Db 202 KNATVTLTSHRTHNLAKVAADILVVAIGRAKFPVTADFKVPGAVVIDVGMNRDENG--- 258
Qy 242 RGYELVGVCEYEASKIAGLITPVGCGVPMTIAMLLSNTLESAGR 287
Db 259 ----KLCDGVDFEAVAPLASHITPVGCGVPMTIAMLLMEQTYQAALR 301

Search completed: July 24, 2004, 06:40:09
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2004, 04:13:15 ; Search time 52 Seconds
(without alignments)
292.394 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIDGKLVAQIREIAVE.....TIAMLLSNTLESAKRIHKEK 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	762	51.3	934	1	CITC DROME
2	739.5	49.8	946	1	CITC YEAST
3	738	49.7	934	1	CITC MOUSE
4	735	49.5	934	1	CITC HUMAN
5	731	49.2	934	1	CITC RAT
6	713.5	48.0	975	1	CLTM YEAST
7	692	46.6	284	1	FOLD STRTR
8	683.5	45.0	283	1	FOLD BACSU
9	681.5	45.9	281	1	FOLD HAEIN
10	679	45.7	287	1	FOLD ECOLI
11	676	45.5	285	1	FOLD PHOPO
12	673	45.3	287	1	FOLD SALTI
13	669	45.0	287	1	FOLD SALTY
14	652.5	43.9	357	1	MTDC DROME
15	629.5	42.4	350	1	MTDC MOUSE
16	623.5	42.0	344	1	MTDC HUMAN
17	608	40.9	286	1	FOLD CHLFP
18	602	40.5	290	1	FOLD HELPY
19	597	40.2	290	1	FOLD HELPY
20	589	39.6	285	1	FOLD BUCAI
21	585.5	39.4	287	1	FOLD CHLMU
22	576.5	38.8	287	1	FOLD CHLPR
23	573.5	38.6	310	1	FOLD RICPR
24	564	38.0	285	1	FOLD BUCAP
25	418	28.1	276	1	FOLD THEAC
26	300	20.2	273	1	FOLD MYCGE
27	286	19.2	269	1	FOLD MYCFN
28	230	15.5	320	1	MTD1 YEAST
29	106	7.1	342	1	OC2D RHIME
30	104	7.0	372	1	DHA2 STAAW
31	103	6.9	372	1	DHA2 STAAW
32	103	6.9	372	1	DHA2 STAAW
33	98	6.6	266	1	AROE HALNI

34 98 371 1 DHA STAEP
35 97.5 230 1 SUMT METIV
36 96 463 1 YELL METUA
37 95 283 1 AROE METTH
38 94 297 1 SVFA BACSU
39 94 378 1 DHA BACSU
40 92.5 373 1 MDHM CHLRE
41 92 502 1 COAT BPTS
42 92 1025 1 YEGO ECOLI
43 91.5 320 1 OCD RHIME
44 91 447 1 MCRB METUA
45 90.5 372 1 DHA BACST

ALIGNMENTS

RESULT 1
CITC DROME STANDARD; PRT; 934 AA.
ID CITC DROME
AC O965E3; Q8T0P2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE [includes: Methyltetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE Methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN PUG OR CG4067
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=99054704; PubMed=9832531;
RA Rong Y.S., Golic K.G.;
RT "Dominant defects in Drosophila eye pigmentation resulting from a
RT euchromatin-heterochromatin fusion gene.";
RL Genetics 150:1551-1566(1998).
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-U., Andrews-Frankoch C., Baldwin D.,
RA Beeson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Fan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RL Science 287:2185-2195(2000).
 [3]
 SEQUENCE OF 569-933 FROM N.A.
 RC STRAIN=Berkely; TISSUE=Head;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
 RA Rubin G.M., Celniker S.E.;
 RA "A *Drosophila* full-length cDNA resource";
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
 CC 5,10-methylenetetrahydrofolate + NADPH.
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 CC formyltetrahydrofolate.
 CC -1- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
 CC phosphate + 10-formyltetrahydrofolate.
 CC -1- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
 CC methionine, histidine, pantothenate, and formyl tRNA-Met.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: Present in all tissues.
 CC -1- MISCELLANEOUS: This trifunctional enzyme consists of two major
 CC domains: a N-terminal part, containing the methylene-THF
 CC dehydrogenase and the methenyl-THF cyclohydrolase activities and a
 CC larger formyl-THF synthetase domain.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, TO OTHER DEHYDROGENASE/
 CC CYCLOHYDROLASE ENZYMES OR DOMAINS.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, STRONG TO OTHER
 CC FORMYLTETRAHYDROFOLATE SYNTHETASE DOMAINS OR ENZYMES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF082097; AAC78847.1; -;
 CC EMBL: AB003658; AAG22140.1; -;
 CC EMBL: AY069146; AAL39291.1; ALT_INIT.
 CC HSP: P11586; I041.
 CC FlyBase: FBgn0020385; pug.
 CC InterPro: IPR000559; Fmtethyd synth.
 CC InterPro: IPR000672; THF_Dhg_Cyh.
 CC Pfam: PF01268; FTHFS; 1.
 CC Pfam: PF00763; THF_DHG_CYH; 1.
 CC Pfam: PF02882; THF_DHG_CYH; 1.
 CC PRINTS: P00085; THFOHROGNASE.
 CC ProDom: PD002300; THF_Dhg_Cyh; 1.
 CC PROSITE: PS00721; FTHFS 1; 1.
 CC PROSITE: PS00722; FTHFS 2; FALSE NEG.
 CC PROSITE: PS00766; THF_DHG_CYH 1; FALSE NEG.
 CC PROSITE: PS00767; THF_DHG_CYH 2; 1.
 CC Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
 KW One-carbon metabolism; ATP-binding; Purine biosynthesis;
 KW Amino-acid biosynthesis; Methionine biosynthesis;
 KW Histidine biosynthesis.
 FT DOMAIN 1 304 METHYLENETETRAHYDROFOLATE DEHYDROGENASE
 FT AND CYCLOHYDROLASE (BY SIMILARITY).
 FT DOMAIN 305 934 FORMYLTETRAHYDROFOLATE SYNTHETASE (BY
 FT SIMILARITY).
 FT NP_BIND 379 386 ATP (BY SIMILARITY).
 FT SEQUENCE 934 AA; 99626 MW; 42CC8A69A4D5345 CRC64;
 SQ

Query Match 51.3%; Score 762; DB 1; Length 934;
 Best Local Similarity 55.2%; Pred. No. 6.3e-51;
 Matches 160; Conservative 38; Mismatches 88; Indels 4; Gaps 3;
 QY 1 AKIDGKLVAKQIREETAIVEIAKM-KDAIGVPGCLAVILVSGSKDSQTYVRNKKACEAV 59
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 4 AKIISGTAFAKSIREEELNRNEVTMSKQLADFPGLRIVQVGGREDSNVYIRMKIKATEI 63
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 60 GIKSYEVNLPEDSSEDEVLKHIATFNSDPVSHGILVOLPL--PHHMDENILNAVSTEKD 117
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 64 GIDAAHVQLPERSITEVELLKINDLNEDPRVHGIIVQMPDCCDTPIDSHRITDAVSPEKD 123
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 118 VDFHPLNIGRLAMQGRDPFFVPCPKGMELHRYGVEIKGRAVVGISNTVGMPAAL 177
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 124 VDLGHTVNEGLAT-GDLGGFLPCTPWGCLLELRSSGVEIAGARAVVILGSKIVGTAAE 182
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 178 LLOKANATVSIIVSNTKKPEEITRQADIVIAAAGVANLVRSWKPKGAATIDVGINPVDD 237
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 183 LLKWNATVTVCHSKIRNLEIITRSADILVVGVAEWKGSWKPGAVVVDGGINVKPD 242
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 238 PESPRGYRLVGDVVCYBASKIAGLITPVGCGVPMITAMLLSNTLSAKR 287
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 QY 243 ASKASGSKLVGDVDAEALQVAGHLTPVPGCGVPMITVAMLMKNTVRSAAAR 292
 DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
 RESULT 2
 ID CLTC_YEAST STANDARD; PRT; 946 AA.
 AC P07245;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
 DE [includes: Methylenetetrahydrofolate dehydrogenase (EC 1.5.1.5);
 DE Methylenetetrahydrofolate cyclohydrolase (EC 3.5.4.9);
 DE Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
 GN ADE3 OR YGR204W OR G7733.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86168166; PubMed=3514599;
 RA Staben C., Rabinowitz J.C.;
 RT "Nucleotide sequence of the *Saccharomyces cerevisiae* ADE3 gene
 RT encoding C1-tetrahydrofolate synthase";
 RL J. Biol. Chem. 261:4629-4637(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=97060019; PubMed=8904340;
 RA Guerreiro P., Barreiros T., Soares H., Cyrne L., Maia e Silva A.,
 RA Rodrigues-Pousada C.;
 RT "Sequencing of a 17.6 kb segment on the right arm of yeast chromosome
 RT VII reveals 12 ORFs, including CCT, ADE3 and TR-I genes, homologues of
 RT the yeast PMT and EFLG genes, of the human and bacterial electron-
 RT transferring flavoproteins (beta-chain) and of the *Escherichia coli*
 RT phosphorine phosphohydrolase, and five new ORFs";
 RL Yeast 12:273-280(1996).
 RN [3]
 RP MUTAGENESIS.
 RX MEDLINE=89247401; PubMed=2541774;
 RA Barlowe C.K., Williams M.E., Rabinowitz J.C., Appling D.R.;
 RT "Site-directed mutagenesis of yeast C1-tetrahydrofolate synthase:
 RT analysis of an overlapping active site in a multifunctional enzyme";
 RL Biochemistry 28:2099-2106(1989).
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
 CC 5,10-methylenetetrahydrofolate + NADPH.
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 CC formyltetrahydrofolate.
 CC -1- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: BC008523; AA08523.1; -
DR MGD: MGI:1342005; Mthcd1.
DR InterPro: IPR000559; Fmtethyd synth.
DR InterPro: IPR000672; THF_Dhg_Cyh.
DR Pfam: PF01268; FTHFS; 1.
DR Pfam: PF00763; THE DHG CYH; C. 1.
DR Pfam: PF02882; THE DHG CYH; C. 1.
DR PRINTS: PR00085; THFDHGNASE.
DR PRODOM: PD002300; THF_Dhg_Cyh; 1.
DR PROSITE: PS00721; FTHFS_1; 1.
DR PROSITE: PS00722; FTHFS_2; 1.
DR PROSITE: PS00766; THE DHG CYH; 1.
DR PROSITE: PS00767; THE DHG CYH; 2; 1.
DR Hydrolase; Ligase; Oxidoreductase; NADP; Multifunctional enzyme;
KW One-carbon metabolism; ATP-binding; Purine biosynthesis;
KW Amino-acid biosynthesis; Methionine biosynthesis;
KW Histidine biosynthesis.
FT INIT MET 0 0
FT DOMAIN 1 304
FT FT
FT DOMAIN 305 934
FT FT
FT NP_BIND 379 386
FT FT
SQ SEQUENCE 934 AA; 101124 MW; DC5C4FE95F59F19 CRC64;

Query Match 49.7%; Score 738; DB 1; Length 934;
Best Local Similarity 49.0%; Pred. No. 4.4e-49;
Matches 145; Conservative 57; Mismatches 90; Indels 4; Gaps 3;

QY 1 AKIIDGLVAKQIREIAVEIAEMKDAI-GVYPLGAVILVGRKDSQTVVRKKACAEAV 59
DB 3 AGLINGKLISQAIRDLKQVTRMQVQFGFGLAILQVGRDSDNLYNKLKAAEEI 62

QY 60 GIKSYEVNLPEDSSEDEVLKHAIFNSDPSVHGILVOLPL--PHMNDENILNAVSIKED 117
DB 63 GIKATHIKLPRTSEVLKVIYSLNEDASVHGFIQVLPDSENSINTEAVINAPEKD 122

QY 118 VDFGFLINTGLRMQGRDFFVPTCPKGCVELLHRYGVBIKGRVAVIGRSNIVGMPAAL 177
DB 123 VDLGTSVSAGKLARGDLNDFCTPCTPKGLEIKAGVQIAGRHAVVWGRSKIVGAPMHD 182

QY 178 LLOKANATYSIVHSNTKKPEETROADIVIAAGVANLVGRGWSIKPGLAIIIDVGINPVDD 237
DB 183 LLLWNNATVTTCHSKTANLDKXNKVDILVATGQPEVMKGEWIKFAGVVIDCGINYPD 242

QY 238 PESPRGYRLVGDVCYBEASIKIAGLITPVGCGVPMTAMLSNTLESAXR-IHKPK 292
DB 243 DTPKNGKRVGVDVAYDEAKERASFITPVGCGVPMTAVLMQSTVESAQRFLOQEK 298

RESULT 4
CLIC_HUMAN STANDARD; PRT; 934 AA.
AC P11586; Q86VC9;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE C-1-tetrahydrofolate synthase, cytoplasmic (C1-THF synthase)
DE [Includes: Methylene-tetrahydrofolate dehydrogenase (EC 1.5.1.5);
DE Methylene-tetrahydrofolate cyclohydrolase (EC 3.5.4.9);
DE Formyltetrahydrofolate synthetase (EC 6.3.4.3)].
GN MTHFD1 OR MTHFD OR MTHFC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-30.
RC TISSUE=Liver;
RX MEDLINE=89034046; PubMed=3053686;

RA Hum D.W., Bell A.W., Rozen R., Mackenzie R.E.;
RT "Primary structure of a human trifunctional enzyme. Isolation of a
RT cDNA encoding methylenetetrahydrofolate dehydrogenase-
RT methylenetetrahydrofolate cyclohydrolase-formyltetrahydrofolate
RT synthetase.";
RL J. Biol. Chem. 263:15946-15950(1988).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=Brain, and Lymph;
RC MEDLINE=22388257; PubMed=12477932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rahas S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-16.
RP TISSUE=Platelet;
RC MEDLINE=22608298; PubMed=12665801;
RX Gevaert K., Goethals M., Martens L., Van Damme J., Staes A.,
RA Thomas G.R., Vandekerckhove J.;
RT "Exploring proteomes and analyzing protein processing by mass
RT spectrometric identification of sorted N-terminal peptides.";
RL Nat. Biotechnol. 21:566-569(2003).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 1-301.
RX MEDLINE=98179934; PubMed=9519408;
RA Allaire M., Li Y., Mackenzie R.E., Cygler M.;
RT "The 3-D structure of a folate-dependent dehydrogenase/cyclohydrolase
RT bifunctional enzyme at 1.5-A resolution.";
RL Structure 6:173-182(1998).
RN [5]
RP VARIANT SPINA BIFIDA HIS-292, AND VARIANT GLN-652.
RX MEDLINE=98272628; PubMed=9611072;
RA Hol F.A., van der Put N.M.J., Geurts M.P.A., Heil S.G.,
RA Trijbels F.J.M., Hamel B.C.J., Mariman E.C.M., Blom H.J.;
RT "Molecular genetic analysis of the gene encoding the trifunctional
RT enzyme MTHFD (methylenetetrahydrofolate-dehydrogenase,
RT methylenetetrahydrofolate-cyclohydrolase, formyltetrahydrofolate
RT synthetase) in patients with neural tube defects.";
RL Clin. Genet. 53:119-125(1998).
CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC 5,10-methylenetetrahydrofolate + NADPH.
CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H2O = 10-
CC formyltetrahydrofolate.
CC -1- CATALYTIC ACTIVITY: ATP + formate + tetrahydrofolate = ADP +
CC phosphate + 10-formyltetrahydrofolate.
CC -1- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
CC methionine, histidine, pantothenate, and formyl tRNA-Met.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DISEASE: Defects in MTHFD1 seem to be associated with elevated
CC risk for neural tube defects (NTD) such as spina bifida.
CC -1- MISCELLANEOUS: This trifunctional enzyme consists of two major
CC domains: a N-terminal part, containing the methylene-THF
CC dehydrogenase and the methenyl-THF cyclohydrolase activities and a
CC larger formyl-THF synthetase domain.

TURN	139	141
HELIX	146	157
TURN	158	158
TURN	163	164
STRAND	166	170
TURN	174	176
HELIX	177	186
TURN	187	188
STRAND	190	194
TURN	196	197
HELIX	201	205
TURN	206	207
STRAND	210	213
TURN	218	219
STRAND	221	221
HELIX	223	225
TURN	228	229
STRAND	231	234
STRAND	238	241
TURN	243	244
TURN	246	247
STRAND	249	252
STRAND	255	255
HELIX	257	260
TURN	261	263
STRAND	266	267
TURN	270	272
TURN	275	276
HELIX	277	294
SEQUENCE	934 AA;	101428 MW; 4AA4678A2BA065E4 CRC64;
Query Match	49.5%;	Score 735; DB 1; Length 934;
Best Local Similarity	49.0%;	Pred. No. 7.5e-49;
Matches 145;	Conservative 56;	Mismatches 91; Indels 4; Gaps 3;
QY	1 AKITDGLVAKOIEETAEVAETAAMKDAI-GVVPGILAVLGVSRKDSQTYYRNKKACEAV	59
DDB	3 AEILINGEISAIQRALKNQTLQKEQVPGETPRLAILQGNRRDSDNYINVKLAEEI	62
QY	60 GKISEVENLPEDSSSEDEVLKHIAIFNSDPNVHGIIVQLPL--PHMNDENILNAVSIEMKD	117
DDB	63 GIKAITHIKLPITTESEVMKYITSLNEDSTVHGFLVQLPLDSENSINTEEVINAPIEKD	122
QY	118 VDGHPHLNIGRANOGRRDFPVCTPTPKCMELLHRHYGVEIKGRRAVTVGRSNIVGMPPAL	177
DDB	123 VDGLTTSINAGRLARGDLNDCFIPTCKGLELIKETGVTPIAGRHAVVVGVRSKIVGAPMD	182
QY	178 LLQRANATVSIVHSNTKKPEBITRQADIVIAAGVANLVRSWIKPGAIIIDVGINPVDD	237
DDB	183 LLLWNNAITVTCCHKTAHLDEVNKGDILLVATGOPENVKGSEWIKPGAIVIDCGINYVPD	242
QY	238 PESPRGYELVDVCYEASKIAGLTTPVPGVGPMTIAMLNSTLESKR-IHKFK	292
DDB	243 DKKPNRGKVGDVAYDEAKERASFITPVGGVGPMTVAMLMOSTVESAKRFLEKF	298
RESULT 5	CITC RAT STANDARD; PRT; 934 AA.	
CITC RAT ID	P27653; Q62808;	
AC	01-AUG-1992 (Rel. 23, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	C-1-tetrahydrofolate synthase, cytoplasmic (Cl-TFH synthase) [includes: Methylentetrahydrofolate dehydrogenase (EC 1.5.1.5); Methylnitrotetrahydrofolate cyclohydrolase (EC 3.5.4.9); Formyltetrahydrofolate synthetase (EC 6.3.4.3)]	
DN	MFHD1 OR MFHPD.	
OS	Rattus norvegicus (Rat).	
OCC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OCC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OR	NCBITaxId=10116;	
OX	[1]	

DR	HSPSP; P11586; IAA1.	
DR	Germline; 138627; -	
DR	S3D; S0000288; MIS1.	
DR	GO; GO:0005739; C:mitochondrion; IDA.	
DR	GO; GO:0004329; F:formate-tetrahydrofolate ligase activity; IDA.	
DR	GO; GO:0004477; F:methylenetetrahydrofolate cyclohydrolase act. .; IDA.	
DR	GO; GO:0004488; F:methylenetetrahydrofolate dehydrogenase (NA. .; IDA.	
DR	GO; GO:0001718; G:conversion of met-tRNA ^f to met-tRNA ⁱ ; IMP.	
DR	GO; GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. .; IMP.	
DR	InterPro; IPR000559; F:mtethyld synth.	
DR	InterPro; IPR000672; THF_Dhg_Cyh.	
DR	Pfam; PF01268; FTHFS; 1.	
DR	Pfam; PF00753; THF_DHG_CYH; 1.	
DR	Pfam; PF02882; THF_DHG_CYH_C; 1.	
DR	PRINTS; PR00085; THFDHGRGNASE.	
DR	ProDom; PD002300; THF_Dhg_Cyh; 1.	
DR	PROSITE; PS00721; FTHFS 1; 1.	
DR	PROSITE; PS00722; FTHFS 2; 1.	
DR	PROSITE; PS00766; THF_DHG_CYH 1; 1.	
DR	PROSITE; PS00767; THF_DHG_CYH 2; 1.	
KW	Hydrolase; ligase; oxidoreductase; NADP; Multifunctional enzyme;	
KW	Mitochondrion; transit peptide; One-carbon metabolism; ATP-binding;	
KW	Purine biosynthesis; Amino-acid biosynthesis; Histidine biosynthesis;	
KW	Methionine biosynthesis.	
FT	TRANSIT 1 34	
FT	CHAIN 35 975	
FT	DOMAIN 35 343	
FT	DOMAIN 344 975	
FT	NP_BIND 408 415	
FT	SEQUENCE 975 AA; 106217 MW; B9421D4F8981531F CRC64;	
Query Match	48.0%; Score 713.5; DB 1; Length 975;	
Best Local Similarity	50.7%; Pred. No. 3.6e-47;	
Matches 145; Conservative	45; Mismatches 95; Indels 1; Gaps 1;	
QY	3 IIDGLVAKQIREIAVEIAKMK-DAIGVVPGLAVILVGSRKDSQTYVRNKKKACEAVGI 61	
Db	ILSGRKLQASIRKANDIEQAIKLKPNFKPTLKIIQVGARPSDSTYYRMLKASKDSNV 96	
QY	62 KSYEVNLPDSSDEVLKHTATFNSDPVHGILLQLPLPHMNDENTINAVSIEKVDVGF 121	
Db	DCIIEKLPAAETVEVELLUKKSIDINDDSIHGILLQLPLRHLDETITINAVDFKKVDVGF 156	
QY	122 HPLNIGRLAMGCRDPFFVPCTPKGCMELLHRYGYEIKGKRAVIGRSIVNGMPAALLQK 181	
Db	157 HRYNAGELAKKGKPYRIPCTPYGCKMLLEEAAHVKLDGKNAVLGRSSIVGNPIASLLKN 216	
QY	182 ANATVSIVHNTKKPEBITROADIVIAAGVANLVGRSWIKFGAAIDVGINPVDDESP 241	
Db	217 ANATVTVCHSHTRNIAEVSQADIVIRACGIPQYVKSDWIEKGAVIDVGINYPDISKK 276	
QY	242 RGVRLGDCVCEEASKIAGLTTPVPGVGVPMTIAMLLSNTLESAR 287	
Db	277 SGQKLGVGDVDESDSVEKTSYITPVPGVGVPMTVAMLVSNVLLAAR 322	
RESULT 7		
FOLD STRTR	STANDARD; PRT; 284 AA.	
ID_FOLD STRTR		
AC	P96050;	
DT	15-JUL-1998 (Rel. 36, Created)	
DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, Last annotation update)	
DE	Fold bifunctional protein [includes: Methylenetetrahydrofolate	
DE	dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase	
GN	(EC 3.5.4.9)].	
OS	Streptococcus thermophilus.	
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OC	Streptococcus.	
OX	NCBI_TaxID=1308;	
RN	[1]	

DE dehydrogenase (EC 1.5.1.5); Methenyltetrahydrofolate cyclohydrolase
 DE (EC 3.5.4.9)].
 GN FOLD OR BSU24310.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RX MEDLINE=9712195; PubMed=8969508;
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
 RA Kobayashi Y.;
 RT "Systematic sequencing of the 23 kb 210 degrees-232 degrees region of
 RT the Bacillus subtilis genome containing the skin element and many
 RT sporulation genes.";
 RL Microbiology 142:3103-3111(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Etian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Hache J., Harwood C.R., Henaut A.,
 RA Hilbert B., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RT Nature 390:249-256(1997).
 RL Nature 390:249-256(1997).
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
 CC 5,10-methylenetetrahydrofolate + NADPH.
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 CC formyltetrahydrofolate.
 CC -!- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
 CC methionine, histidine, pantothenate, and formyl tRNA-Met.
 CC -!- SUBUNIT: Homodimer (by similarity).
 CC -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
 CC DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D84432; BAA12572.1; -
 CC EMBL: 299116; CAB14362.1; -
 CC EMBL: E69626; E69626.
 CC HSP: P11586; I441.

DR Subtilist; BG11711; fold.
 DR Interpro; IPR000672; THF DHG Cyt.
 DR Pfam; PF00763; THF DHG Cyt; 1.
 DR Pfam; PF02882; THF DHG Cyt; 1.
 DR PRINTS; PR00085; THFDHGRNASE.
 DR ProDom; PD002300; THF DHG Cyt; 1.
 DR PROSITE; PS00766; THF DHG Cyt; 1.
 DR PROSITE; PS00767; THF DHG Cyt; 1.
 KW Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
 KW Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
 KW Histidine biosynthesis; Hydrolase; Complete proteome.
 SQ SEQUENCE 283 AA; 30685 MW; E004E0AC40A30889 CRC64;
 Query Match 46.0%; Score 683.5; DB 1; Length 283;
 Best Local Similarity 49.1%; Pred. No. 1.6e-45;
 Matches 141; Conservative 50; Mismatches 87; Indels 9; Gaps 4;
 QY 1 AKIDGKLVAKQIREETAVETAKMKDAIGVPGGLAVILVGRKDSOTYYRNKKKCAEAVG 60
 DB 3 ATIDGKETAREKREQLAKEVELKKQ-GVTPGLAVILIGDDPASHYVRGKKAAETMG 61
 QY 61 IKSEVNLPESSSEDEVKHIATFNSDPSVHGILVOLPLPHMNDENILNVAISIEKVDVG 120
 DB 62 MNFKLDQFSSLSAEALLSIIDQYNQDPEPHGILVOLPLPDHISEKAVIERISPKDVG 121
 QY 121 FHPNLGRAMQGRDPFFVCTPKGWEHLHRYGVKIKRAVIGRSNIVGMPALLQ 180
 DB 122 FHPNLGKML-GEDT-FLPCTPHGIVELKKTNIDLSKEVVVGRSNIVGKPVQQLJL 179
 QY 181 KANATVISVHSNTKPKPEITRQADIVIAVGVANLVGRSGVIAIIVGINPVDPPES 240
 DB 180 NENATVYCHSRNTENITEHTKADILVAVGRANFISADQIKEGAVVIDGVNRLNG-- 237
 QY 241 PRGVRLGVGVCEASKIALITPVGGVGPMTIAMLLSNTLESAR 287
 DB 238 ----KLGGDVEFEKAKASFITPVGGVGPMTITLAHNTVKSAR 280
 RESULT 9
 FOLD HAEIN STANDARD; PRT; 281 AA.
 AC P44313;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE FOLD bifunctional protein [includes: Methylenetetrahydrofolate
 DE dehydrogenase (EC 1.5.1.5); Methylenetetrahydrofolate cyclohydrolase
 DE (EC 3.5.4.9)].
 GN FOLD OR H10609.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kervatage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Frichman J.L., Fuhrmann J.B., Geoghegan N.S.M.,
 RA Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd".
 RL Science 269:496-512(1995).
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
 CC 5,10-methylenetetrahydrofolate + NADPH.
 CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 CC formyltetrahydrofolate.

CC -!- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
CC methionine, histidine, pantothenate, and formyl tRNA-Met.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U32743; AAC22268.1; -
DR PIR; A64081; A64081.
DR HSSP; P11586; 1A41.
DR TIGR; H10609; -
DR InterPro; IPR000672; THF_DHG_CyH.
DR Pfam; PF00763; THF_DHG_CyH; 1.
DR Pfam; PF02882; THF_DHG_CyH; 1.
DR PRINTS; PR00085; THFDRGNASE.
DR PRODOM; PD002300; THF_DHG_CyH; 1.
DR PROSITE; PS00766; THF_DHG_CyH; 1.
DR PROSITE; PS00767; THF_DHG_CyH; 1.
KW Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW Histidine biosynthesis; Hydroxylase; Complete proteome.
FT INIT MET 0 BY SIMILARITY
SQ SEQUENCE 281 AA; 30329 MW; 932ABAB9778CF51C CRC64;

Query Match 45.7%; Score 681.5; DB 1; Length 281;
Best Local Similarity 51.1%; Pred. No. 2.3e-45;
Matches 144; Conservative 44; Mismatches 85; Indels 9; Gaps 2;

QY 1 AKITDKLVAQIREETIAVEIAKMKDAIGVVGAVILVSGSRKDSQTVVRNKKACEAVG 60
DB 2 AKISGTELPYQKANKLADKTHVIEQGRAPGAVILVSGDRPASQYVGNKRKSCSEVG 61
QY 61 IKSYEVNLPEDSSEDEVLKHIATFNSPDSVHGILVOLPLPHHMDENILNAVSTKDVGD 120
DB 62 ILKSYDLPETTTQNELLAIIQDLNADKNIDGILVOLPLPQINAEAIERIDPKDVGD 121
QY 121 FHLNIGRLANQGRDPFPVCPGKCMELLHRYGVETKGAIVGNSNIVGMPAAALLQ 180
DB 122 FHPYNVGRLCQ--RIPTLRACCTPYGWMKLETTGIDLHGKHAIVGASNIYGRPMSELL 179
QY 181 KANATVSVHSNTKKPEEITRQADIVIAAGVANLVGRSGWIKPGAAIIDVGINPVDPPES 240
DB 180 LAGATVTVTRFTKLEHNHVRQADILVAVGKPNLISGDWIKESAVVIDVGINRVDG--- 236
QY 241 PRGVLVGDVGYEAEASKIAGLITPVPGVGPMPTIAMLSNTL 282
DB 237 ---KLGVDFEFDKAAEAAVITPVPGVGPMPTIAMLSNTL 274

RESULT 10
FOLD_ECOLI
ID FOLD_ECOLI STANDARD; PRT; 287 AA.
AC P24186; P77132;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Fold bifunctional protein [Includes: Methyltetrahydrofolate
DE dehydrogenase (EC 1.5.1.5); Methyltetrahydrofolate cyclohydrolase
DE (EC 3.5.4.9)].
GN FOLD OR ADS OR B0529.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-35.

RX MEDLINE=92084696; PubMed=1748668;
RA D'Arì L., Rabinowitz J.C.;
RT "Purification, characterization, cloning, and amino acid sequence of
RT the bifunctional enzyme 5,10-methylenetetrahydrofolate
RT dehydrogenase/5,10-methylenetetrahydrofolate cyclohydrolase from
RT Escherichia coli.";
RL J. Biol. Chem. 266:23953-23958 (1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Yonetani Y., Sanpei G., Mizobuchi K.;
RL Submitted (XXK-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Gosslen D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RT "Sequence of minutes 4-25 of Escherichia coli.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC 5,10-methylenetetrahydrofolate + NADPH.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
CC formyltetrahydrofolate.
CC -!- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
CC methionine, histidine, pantothenate, and formyl tRNA-Met.
CC -!- SUBUNIT: Homodimer.
CC -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP.
CC -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; M74789; AAA23803.1; -
DR EMBL; D10588; BAA01445.1; -
DR EMBL; AF000159; AAC73631.1; -
DR EMBL; U82664; AAB40282.1; -
DR PIR; H64784; J80662.
DR PDB; 1BOA; 29-JUN-99.
DR EcoGene; EG10328; fold.
DR InterPro; IPR000672; THF_DHG_CyH.
DR Pfam; PF00763; THF_DHG_CyH; 1.
DR Pfam; PF02882; THF_DHG_CyH; 1.
DR PRINTS; PR00085; THFDRGNASE.
DR PRODOM; PD002300; THF_DHG_CyH; 1.
DR PROSITE; PS00766; THF_DHG_CyH; 1.
DR PROSITE; PS00767; THF_DHG_CyH; 1.
KW Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
KW Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
KW Histidine biosynthesis; Hydroxylase; Complete proteome; 3D-structure.
FT INIT MET 0
FT CONFLICT 46 S -> L (IN REF. 4).
FT CONFLICT 199 V -> L (IN REF. 1).
SQ SEQUENCE 287 AA; 30912 MW; 163F6BDD9024514E CRC64;

Query Match 45.7%; Score 679; DB 1; Length 287;
Best Local Similarity 48.6%; Pred. No. 3.6e-45;
Matches 141; Conservative 50; Mismatches 89; Indels 10; Gaps 4;

QY 1 AKIIDGKLVAKQIREETIAVEIAKMDAIGV-VFGLAVILVGSKDSQTVYRNKKKACEAV 59
 DB 2 AKIIDGKTIQAQVRSEVQAKV-QAR-AGLRAPGLAVLVGSPASQIYVASKRACEEV 60
 QY 60 GIKSEYVNLPEDESSEDEVLKHIAIFNSDPSVHGILVQLPLPHMNDENILNAVSIKQVD 119
 DB 61 GFVRSYDLPEETTSEAELELIDTLNADNTIDGILVQLPLPAGIDNVKVLRIHPDKVD 120
 QY 120 GFHPLNIGRLAQOGRDPFPVCTPKGCMELHRYGVEIKGRAVIVGRSNIYVGMPPAALL 179
 DB 121 GFHPYNGRLCO--RAPLRPCTPRGIVILLERYNIDTGLNVLGASNIYVGRPMSEL 178
 QY 180 QKANATVSIHVSNTKPEETROQADIVIAAVGVANLVGRSWIKPGAIIIDVGINPVDDE 239
 DB 179 LLACGTTTTHRTKRLRHVENADLLIIVAVGKPGFIPGDWIKGAIIVDVGINRLNG- 237
 QY 240 SPRGYRLVGVCEEAASKIAGLITPVPGGVGPMTIAMLSNTLESASAKRIH 289
 DB 238 -----KVGVDVFEAKRASITPVPGGVGPMVATLIENTILOACVEYH 282

RESULT 11
 FOLD_PROPO STANDARD; PRT; 285 AA.
 AC PS1696;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FOLD bifunctional protein [Includes: Methylenetetrahydrofolate
 dehydrogenase (EC 1.5.1.5); Methylenetetrahydrofolate cyclohydrolase
 (EC 3.5.4.9)].
 GN FOLD
 OS Photobacterium phosphoreum.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Photobacterium.
 OX NCBI_TaxID=659;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 844;
 RA Pawelek P.D., Meija N., Mackenzie R.E.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
 5,10-methylenetetrahydrofolate + NADPH.
 CC -1- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
 formyltetrahydrofolate.
 CC -1- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
 methionine, histidine, pantothenate, and formyl tRNA-Met.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; U34207; AAA98507.1; --
 CC PIR; S71924; S71924.
 CC HSSP; P11586; I041.
 CC InterPro; IPR000672; THF_Dhg_Cyh.
 CC Pfam; PF00763; THE_DHG_CYH; 1.
 CC Pfam; PF02882; THE_DHG_CYH; 1.
 CC PRINTS; PR00085; THFDHGRNASE.
 CC ProDom; PD002300; THE_Dhg_Cyh; 1.
 CC PROSITE; PS00766; THE_DHG_CYH; 1.
 CC PROSITE; PS00767; THE_DHG_CYH; 1.
 CC Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
 KW Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
 KW Histidine biosynthesis; Hydrolase.

SQ SEQUENCE 285 AA; 30775 MW; 9EB1FC954EAA4396 CRC64;
 Query Match 45.5%; Score 676; DB 1; Length 285;
 Best Local Similarity 48.8%; Pred. No. 6.le-45;
 Matches 141; Conservative 46; Mismatches 94; Indels 8; Gaps 2;
 QY 1 AKIIDGKLVAKQIREETIAVEIAKMDAIGVVPGLAVILVGSKDSQTVYRNKKKACEAV 60
 DB 3 AQIIDGKIISQTVRQVEAARVARTDAGLRAPGLAVLVGQDPASQIYVSKRACEEV 62
 QY 61 IKSEYVNLPEDESSEDEVLKHIAIFNSDPSVHGILVQLPLPHMNDENILNAVSIKQVD 120
 DB 63 FIKSFOLFPSASEQQQLDLIDELNQDPTMDGLIVQLPLPAGIDNTRILERIDEKVD 122
 QY 121 FHLNIGRLAQOGRDPFPVCTPKGCMELHRYGVEIKGRAVIVGRSNIYVGMPPAALL 180
 DB 123 FHPYNGRLCO--RIPKLRSCPTKGIITLLERYNIEVRGHAVIVGASNIYVGRPMTE 180
 QY 181 KANATVSIHVSNTKPEETROQADIVIAAVGVANLVGRSWIKPGAIIIDVGINPVDDE 240
 DB 181 LAGATTTTCHRTQDLEGHIRQADILVAVGKFNFPGGWIKGATVIDVGINRLNG-- 238
 QY 241 PRGYRLVGVCEEAASKIAGLITPVPGGVGPMTIAMLSNTLESASAKRIH 289
 DB 239 ----KLCGDVEFDVACORAKYITPVPGGVGPMVATLIENTILOACVEYH 283

RESULT 12
 FOLD_SALTI STANDARD; PRT; 287 AA.
 AC Q60006;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE FOLD bifunctional protein [Includes: Methylenetetrahydrofolate
 dehydrogenase (EC 1.5.1.5); Methylenetetrahydrofolate cyclohydrolase
 (EC 3.5.4.9)].
 GN FOLD OR STY0588 OR T2321.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=STY4;
 EX MEDLINE=94327079; PubMed=8050713;
 RA Rossolini G.M., Muscas P., Chiesurin A., Satta G.;
 RT "fima-fold genes linkage in Salmonella identifies a putative
 RT functional site of chromosomal rearrangement in the enterobacterial
 RT genome.";
 RL FEMS Microbiol. Lett. 119:321-328(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Church C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Tagelis K.,
 RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2


```
RT and Ctl8";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC 5,10-methylenetetrahydrofolate + NADPH.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
CC formyltetrahydrofolate.
CC -!- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
CC methionine, histidine, pantothenate, and formyl tRNA-Met.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74603; CAA52683.1; -.
CC EMBL; AL627267; CAD05024.1; -.
CC EMBL; AE016841; AAC06915.1; -.
CC PIR; S36633; S36633.
CC HSP; P11586; 1A41.
CC InterPro; IPR000672; THE_DHG_CyH.
CC Pfam; PF00763; THE_DHG_CyH; 1.
CC Pfam; PF02882; THE_DHG_CyH_C; 1.
CC PRINTS; PR00085; THEDHGRNASE.
CC PRODOM; PD002300; THE_DHG_CyH; 1.
CC PROSITE; PS00766; THE_DHG_CyH_1; 1.
CC PROSITE; PS00767; THE_DHG_CyH_2; 1.
CC KW Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
CC Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
CC Histidine biosynthesis; Hydrolyase; Complete proteome.
CC INIT_MET 0 BY SIMILARITY.
CC FT INIT_MET 0 BY SIMILARITY.
CC SQ SEQUENCE 287 AA; 30693 MW; D44E18D51701B14D CRC64;

Query Match 45.3%; Score 673; DB 1; Length 287;
Best Local Similarity 47.9%; Pred. No. 1e-44;
Matches 139; Conservative 52; Mismatches 89; Indels 10; Gaps 4;

QY 1 AKIIDKLVAKQIREIAVEIAKMKDAIGV-VPLGAVILVSGRSDSQTYVNRKKACEAV 59
DB 1 AKIIDKLVAKQIREIAVEIAKMKDAIGV-VPLGAVILVSGRSDSQTYVNRKKACEAV 60
QY 60 GIKSYENVLPEDSSDEVILKHIATFNSDPSVHGILVQLPLPHMNDENILNAVSIKDV 119
DB 61 GFVSHSYDLPETTSEAEALLALIDLINADNTIDGILVQLPLPAGIDNVKLERIAPDKVD 120
QY 120 GFHPLNIGRLAQGRDPFFVCTPKGMELHRYGVEIKGRAVVGSRNIVGMPAALLL 179
DB 121 GFHPYNGVRLCQ--RAPRLRPTCPRGIVTLLERYNIDTYGLNAVVGASNVGRPMSEL 178
QY 180 QKANATVSIHNSNTKKPEEITRQADIVIAAVGVANLVRSWIKPGAAIIDVGINPVDPE 239
DB 179 LLAGCTTTVTHRTKDLRHHVHADLLIVAVGKPGFIPGEWIKEGAIVDVGINLENG- 237
QY 240 SPRGYRLGVDCVCEASKIAGLITVPVGGVGPMTIAMLLSNTLESKRHH 289
DB 238 ----KVGVDVDFDEAARASVITPVGPGVPMVTATIENTLQACIEYH 282

RESULT 13
FOLD SALTY STANDARD; PRT; 287 AA.
AC P58688;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fold bifunctional protein [includes: Methylenetetrahydrofolate
DE dehydrogenase (EC 1.5.1.5); Methylenetetrahydrofolate cyclohydrolase
```

```
DE (EC 3.5.4.9)].
GN FOLD OR STM0542.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11577609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + NADP(+) =
CC 5,10-methylenetetrahydrofolate + NADPH.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + H(2)O = 10-
CC formyltetrahydrofolate.
CC -!- PATHWAY: Necessary for the biosynthesis of purines, thymidylate,
CC methionine, histidine, pantothenate, and formyl tRNA-Met.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- MISCELLANEOUS: THIS ENZYME IS SPECIFIC FOR NADP (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER DEHYDROGENASE/CYCLOHYDROLASE ENZYMES OR
CC DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE008721; AAL19496.1; -.
CC DR StyGene; SG????; fold.
CC DR InterPro; IPR000672; THE_DHG_CyH.
CC Pfam; PF00763; THE_DHG_CyH; 1.
CC Pfam; PF02882; THE_DHG_CyH_C; 1.
CC PRINTS; PR00085; THEDHGRNASE.
CC PRODOM; PD002300; THE_DHG_CyH; 1.
CC PROSITE; PS00766; THE_DHG_CyH_1; 1.
CC PROSITE; PS00767; THE_DHG_CyH_2; 1.
CC KW Multifunctional enzyme; One-carbon metabolism; Oxidoreductase; NADP;
CC Purine biosynthesis; Amino-acid biosynthesis; Methionine biosynthesis;
CC Histidine biosynthesis; Hydrolyase; Complete proteome.
CC INIT_MET 0 BY SIMILARITY.
CC SQ SEQUENCE 287 AA; 30712 MW; 26A97583C36CE103 CRC64;

Query Match 45.0%; Score 669; DB 1; Length 287;
Best Local Similarity 47.9%; Pred. No. 2.1e-44;
Matches 139; Conservative 51; Mismatches 90; Indels 10; Gaps 4;

QY 1 AKIIDKLVAKQIREIAVEIAKMKDAIGV-VPLGAVILVSGRSDSQTYVNRKKACEAV 59
DB 2 AKIIDKLVAKQIREIAVEIAKMKDAIGV-VPLGAVILVSGRSDSQTYVNRKKACEAV 60
QY 60 GIKSYENVLPEDSSDEVILKHIATFNSDPSVHGILVQLPLPHMNDENILNAVSIKDV 119
DB 61 GFVSHSYDLPETTSEAEALLALIDLINADNTIDGILVQLPLPAGIDNVKLERIAPDKVD 120
QY 120 GFHPLNIGRLAQGRDPFFVCTPKGMELHRYGVEIKGRAVVGSRNIVGMPAALLL 179
DB 121 GFHPYNGVRLCQ--RAPRLRPTCPRGIVTLLERYNIDTYGLNAVVGASNVGRPMSEL 178
QY 180 QKANATVSIHNSNTKKPEEITRQADIVIAAVGVANLVRSWIKPGAAIIDVGINPVDPE 239
DB 179 LLAGCTTTVTHRTKDLRHHVHADLLIVAVGKPGFIPGEWIKEGAIVDVGINLENG- 237
QY 240 SPRGYRLGVDCVCEASKIAGLITVPVGGVGPMTIAMLLSNTLESKRHH 289
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:47:07 ; Search time 3406 Seconds
(without alignments)
2560.116 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIIDGKLVAKQIREIAVE.....TIAMLLSNTLESARKRIHKPK 292

Scoring table:

BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10723061/runat_21072004_090401_7121/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061.cgn_1_1_6425@runat_21072004_090401_7121 -NCPH=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hct:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1143	76.9	892	14	CD438219 EL01N0510
2	1024	68.9	1087	11	AY106093 Zea mays
3	997	67.1	737	13	BQ123966 EST609542
4	992	66.8	860	13	BQ165274 EST611143
5	985	66.3	621	14	CD057603 HO15K04S
6	982	66.1	832	13	BQ804363 WHE3353 F
7	982	66.1	838	14	CD574828 UCRPT01 0
8	964	64.9	605	10	B3359332 DGL 40 F1
9	945	63.6	633	12	BG415373 HVSMEK00
10	934	62.9	870	14	CK298478 EST761192
11	925	62.2	586	13	CA008897 HUI2H11r
12	924	62.2	595	13	BQ609202 BRY 5126
13	912	61.4	762	13	BQ295278 WHE2868 B
14	908	61.1	708	13	BQ788577 WHE34151 D
15	861	57.9	797	14	CS686156 Bn01b_020
16	854	57.5	696	10	AW776824 EST335889
17	850	57.2	789	12	BJ197939 EST339674
18	841	56.6	648	13	BQ123426 EST6609002
19	840	56.5	655	13	BU546290 GM880009B
20	840	56.5	726	10	AW776355 EST335420
21	836	56.3	859	14	CK294984 EST757698
22	835	56.2	754	13	CA124209 SCQGLR106
23	823	55.4	762	14	CF205415 RR809151
24	818	55.0	709	10	BE202855 EST1402877
25	818	55.0	902	11	AY112351 Zea mays
26	817	55.0	716	12	BM448325 DSA024812
27	816	54.9	849	14	CB974765 CAB30005
28	815	54.8	691	14	CD813389 BN15_019G
29	815	54.8	708	10	BE038548 AB01A07 A
30	814	54.8	604	13	BQ394217 QGF6K04_Y
31	811.5	54.6	877	12	BM817306 HCL08B02_
32	811	54.6	634	12	BJ230917 BJ230917
33	811	54.6	644	12	BJ230946 BJ230946
34	811	54.6	724	12	BJ242965 BJ242965
35	807	54.3	707	14	CD900176 G174.114L
36	806	54.2	843	10	AW883356 HVSMEG001
37	803	54.0	700	9	AL506837 AL506837
38	801	53.9	691	12	BJ226005 BJ226005
39	799	53.8	717	14	CD875624 AZ03.106H
40	798	53.7	716	13	BQ579696 WHE2973 E
41	797	53.6	739	12	BJ576118 BJ576118
42	786	52.9	754	14	CD900177 G174.114L
43	782.5	52.7	699	14	CD813202 BN15_001N
44	777	52.3	720	12	BJ250385 BJ250385
45	777	52.3	722	12	BJ244554 BJ244554

ALIGNMENTS

RESULT 1
CD438219
LOCUS EL01N0510G03.b Endosperm_5 Zea mays cDNA, mRNA linear EST 03-JUN-2003
DEFINITION CD438219
ACCESSION CD438219
VERSION CD438219.1 GI:31353862
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 892)

AUTHORS Lai, J., Dey, N., Kim, C.S., Becraft, P., Larkins, B., Linton, E. and
Messing, J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished (2002)
COMMENT Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.

FEATURES source
1. .892 Location/Qualifiers
/organism="Zea mays"
/mol_type="mRNA"
/cutoff="W22"
/db_xref="taxon:4577"
/tissue_type="Endosperm of 7-23DAP"
/clone_lib="Endosperm 5"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
Pred. No.: 6,81e-121 Length: 892
Score: 1143.00 Matches: 221
Percent Similarity: 95.80% Conservative: 7
Best Local Similarity: 92.86% Mismatches: 10
Query Match: 76.92% Indels: 0
DB: 14 Gaps: 0

US-10-723-061-18 (1-292) x CD438219 (1-892)

QY 55 AlaCysGluAlaValGlyLeuYsSerTyGluValAsnLeuProGluAspSerSerGlu 74
Db 10 GCCTGGCAGCGGTCGGTAAAGTCGTACGAGGTCCTTTCCCTGAGACAGCTCTGAG 69

QY 75 AspGluValLeuYsHisTleAlaThrPheAsnSerAspProSerValHisGlyLeu 94
Db 70 GAGGAGGTTATCAAGCACATAGGAGTTTCAACAGTACGCCCTCGCTGGCATCTG 129

QY 95 ValGluLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAlaValSerIle 114
Db 130 GTTCAGTTGCCCTTACCAAGGATATGAACGATGAGCAATCTTGAATCGGTCAGTATT 189

QY 115 GluYsAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArg 134
Db 190 GAAAGGATGTTGATGGCTTTCATCGCTGAACATCGGACGCGCTGCAATGCAAGCCGA 249

QY 135 AspProPhePheValProCysThrProYsGlyCysMetGluLeuLeuHisArgTyGly 154
Db 250 GATCCGTTCTTTGTACCTTGCACCCGAAAGGATGATGAGTGTCTACCGACACCGA 309

QY 155 ValGluIleYsGlyYsArgAlaValIleGlyArgSerAsnIleValGlyMetPro 174
Db 310 GTTGAATTAAGGGAAGAGAGCTGTTGTATCGACGACGAGCAATATTGTTGGAACGCC 369

QY 175 AlaAlaLeuLeuGlnYsAlaAsnAlaThrValSerIleValHisSerAsnThrYs 194
Db 370 GCTGCATTTGTTGCAAAAGCAATGCACTGTGAGCATTTGTTCAATTCACAAACCAAG 429

QY 195 LysProGluLeuIleThrArgGlnAlaAspIleValIleAlaValGlyValAlaAsn 214
Db 430 AACCTGACCAATACGACGAGGAGGATATGATCATCGACGCTGTTGGAGTCCCAAC 489

QY 215 LeuValArgGlySerTrpIleYsProGlyAlaAlaIleIleAspValGlyIleAsnPro 234
Db 490 TTGGTCAGAGGGGANTGGATAAAGCTCGAGCTACTATTATTGATGTTGGCATCAATCCA 549

QY 235 ValAspAspProGluSerProArgGlyTyArgLeuValGlyAspValCysTyGluGlu 254
Db 550 GTTGATGACCCAGAGAGCCCCCGGGGTACAGGCTCGTCGGGGAGCTCTGCTACAGGAA 609

QY 255 AlaSerYsIleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIle 274
Db 610 GCCTCCAAAGTCGCGGAGGCATCACGCGGTCGCCGCGCGGCGGATGACGATC 669

QY 275 AlaMetLeuSerAsnThrLeuGluSerAlaYsArgIleHisYsPheYs 292
Db 670 CGGATGCTTCTGTCGACACGCTCGAGTCGCGCTAGACGATCCAGGATTCAG 723

RESULT 2

AY106093 1087 bp mRNA linear HTC 16-OCT-2002
LOCUS Zea mays PCOL12665 mRNA sequence.
ACCESSION AY106093
VERSION AY106093.1 GI:21209171
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 1087)
AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

TITLE

JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1087)
AUTHORS Coe, E.H.
JOURNAL Direct Submission

COMMENT

Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSI, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.

FEATURES

Location/Qualifiers
1. .1087
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:836563"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

ORIGIN

Alignment Scores:
Pred. No.: 4,93e-107 Length: 1087
Score: 1024.00 Matches: 191
Percent Similarity: 81.53% Conservative: 43
Best Local Similarity: 66.55% Mismatches: 53
Query Match: 68.91% Indels: 0
DB: 11 Gaps: 0

US-10-723-061-18 (1-292) x AY106093 (1-1087)

QY 1 AlaYsIleIleAspGlyYsLeuValAlaYsGlnIleArgGluGluIleAlaValGlu 20
Db 32 GCACAGATCATCGATGCAAGGCATCGCGCGGCGGAGTGCGCGCGAT 91

QY 21 IleAlaYsMetYsAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 92 GTGGCGCGCTCTCTGTCGCGCCACAGACATCGTGGCGGGGCTGGCCGTGTCATGTGGGG 151

QY	41	SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly	60
Db	152	AGCAGAAGGACTCCGACAGCAGTACGTGAACATGAAGCGCAAGCGTCGCCGAGTCGCG	211
QY	61	IleLysSerTyrGluValAspLeuProGluAspSerSerGluAspGluValLeuLysHis	80
Db	212	ATCTGCTCCATCGACGTCGACCTCCGGAGGACATCTCCGAGACCGGCTGTCGCCGAG	271
QY	81	IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro	100
Db	272	GTTTCATCGCCTCAACGCTGACCGCGCGTCACGGGATCCTTCTGCAGGTTCCACTTC	331
QY	101	HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly	120
Db	332	AGCATATCTCAACGAGAGAGATCTGACGAGATTTCATTCAGAAAGATGTGATGCG	391
QY	121	PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro	140
Db	392	TTCATCTCTCAACATTGGCAATGAAGCCAGAGACCACTGTTCGTACCA	451
QY	141	CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLysGlyLys	160
Db	452	TGTACGCCAAAGGGGTGATGAGCTCTTGTCAAGGAGCGGATCACTGTTAAAGTAAG	511
QY	161	ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln	180
Db	512	CGGCGAGTTGTGGTTCGCGAGCAACATCGTCGGCTACTGTATCTGCTCCTCTCG	571
QY	181	LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluLeuThr	200
Db	572	AAGGCGAGATCGCACCGTATCTGTTGTGCATCGCGGACCCCTGATCCTGAAAGCAT	631
QY	201	ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp	220
Db	632	CGCGAAGCTGACATAGTTCATCGGCGAGCTGGGCGAGCTATGATGATCAAGGTGAC	691
QY	221	IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer	240
Db	692	ATCAAGCCAGGTGTCGCGTTCATGATGTCGGGACGAATCCATCGATGATCACTAC	751
QY	241	ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly	260
Db	752	AGTCCGGGTACCGGCTGTTCGGGATGTGATTTCGACGGCGGAGCAAGGTGTCTGG	811
QY	261	LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn	280
Db	812	TACCTGACTCCGTTCCCGAGGCGTTGGCCCAATGACGTGGCAATGTCTGTGAAGAAC	871
QY	281	ThrLeuGluSerAlaLysArg	287
Db	872	ACGGTGGATGGGCAAGCGG	892
RESULT	3		
LOCUS	BQ123966		
DEFINITION	EST609542 GLSD Medicago truncatula cDNA clone pGLSD-33N16, mRNA	737 bp	linear
ACCESSION	BQ123966		EST 17-APR-2000
VERSION	BQ123966.1		
KEYWORDS	GI:20175928		
SOURCE	EST.		
ORGANISM	Medicago truncatula (barrel medic)		
REFERENCE	1 (bases 1 to 737)		
AUTHORS	Grusak, M.A., Samac, D., Town, C.D., Van Aken, S., Utterback, T., Cheung, F. and Fraser, C.M.		
TITLE	ESTs from late stage developing seeds of Medicago truncatula		
JOURNAL	unpublished (2002)		
COMMENT	Contact: Grusak, M.A. USDA/ARS Children's Nutrition Research Center		

Baylor College of Medicine
 1100 Bates Street, Houston, TX 77030-2600, USA
 Tel: 713 798 7044
 Fax: 713 798 7078
 Email: mgrusak@bcm.tmc.edu
 TIGR sequence name: MRPAX90TK More information is available at:
www.medicago.org
 Seq primer: Skmod (CTA GAA CTA gtg GAT CC).
 Location/Qualifiers
 1..737
 /organism="Medicago truncatula"
 /mol_type="mRNA"
 /cultivar="A17"
 /db_xref="taxon:3880"
 /clone="pGLSD-33N16"
 /tissue_type="Immature seeds"
 /dev_stage="25 to 35 days after pollination"
 /lab_host="XL0LR"
 /clone_lib="GLSD"
 /note="Vector: pluscript SK-; Site_1: EcoRI; Site_2:
 XhoI; Immature seeds, collected from pods ranging in age
 from 25 to 35 days after pollination, were harvested from
 greenhouse-grown plants. Seed were removed and separated
 from pod walls and immediately frozen in liquid nitrogen.
 Seeds throughout the age range were pooled for mRNA
 extraction. cDNA was prepared from polyA+ enriched RNA.
 The cDNA was directionally ligated into the Unizap XR
 vector from Stratagene and packaged using Gigapack III
 Gold packaging extracts. Plasmids containing cDNA inserts
 were excised from the recombinant lambda-zap phage using
 Ex-assist helper phage and propagated in XL0LR cells."

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="Al7"
/db_xref="taxon:3880"
/clone="pGLSD-32N16"
/tissue_type="Immature seeds"
/dev_stage="25 to 35 days after pollination"
/lab_host="XLOLR"
/clone_lib="GLSD"
/note="vector: pluscript SK-; Site_1: EcoRI; Site_2:
XhoI; Immature seeds, collected from pods ranging in age
from 25 to 35 days after pollination, were harvested from
greenhouse-grown plants. Seed were removed and separated
from pod walls and immediately frozen in liquid nitrogen.
Seeds throughout the age range were pooled for mRNA
extraction. cDNA was prepared from polyA+ enriched RNA.
The cDNA was directionally ligated into the Unizap XR
vector from Stratagene and packaged using Gigapack III
gold packaging extracts. Plasmids containing cDNA inserts
were excised from the recombinant lambda-zap phage using
Ex-assist helper phage and propagated in XLOLR cells."

```

ORIGIN

Alignment Scores:		
Pred. No.:	3.61e-104	Length:
Score:	997.00	Matches:
Percent Similarity:	89.91%	Conservative:
Best Local Similarity:	82.89%	Mismatches:
Query Match:	67.09%	Indels:
DB:	13	Gaps:
		737
		189
		23
		0
		0
US-10-723-061-18 (1-292)	x	BO123966 (1-737)

```

US-10-723-061-18 (1-292) x BQ123966 (1-737)
Qy      65  GluValAsnLeuProGluUAspSerSerGluUAspValLeuIleuIysHisIleAlaThrPhe 84
Db      -3  GAAGTAAATTTTCCCGAGGATCTTCGCAAGAAGAAGTTTTTGAACCATATTTGCAGGATT 62
Qy      85  AsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHisMetAsn 104
Db      63  AATGATGATCCTTCGGTTCA TGCGCATTC TGTTCAGCTACCTTTACCATCTCATATGAAT 122
Qy      105 AspGluAsnIleLeuAsnAlaValSerIleGluIysAspValAspGlyPheHisProLeu 124
Db      123 GAACAAATGTGTTCGAATGCTGTAGAAATCGAGAAGCATGTAGATGGTTTTCATCCATTG 182
Qy      125 AsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCysThrProLys 144
Db      183 AATATTGTCGGCTTGCTATCGAGGGAGAGACCCCTGTTTGTTCCTGTACACCAAAG 242
Qy      145 GlyCysMetGluLeuLeuHisArgTyrGlyValgluIleIysGlyIysArgAlaValVal 164
Db      243 GGATGCATAGAACTACTCCAGATATGCGCTTTCTATCAAGAAGAAGGGCTGTGTG 302
Qy      165 IleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnIysAlaAsnAla 184
Db      303 ATTGGCCGTAGTAATATTGTTGGAAATCCAGCTCCTCTCTTGTTCAAAGGGAAGATGCT 362
Qy      185 ThrValSerIleValHisSerAsnThrIysIysProgluIleThrArgGlnAlaAsp 204
Db      363 ACTGTCAGTATTGTTTCATTCAGAACTACTAACCTGAAAGAGATCATTAGACAAAGCAGAT 422
Qy      205 IleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerThrIleIysProGly 224

```

RESULT 3	
BQ123966	
LOCUS	737 bp mRNA linear EST 17-APR-2000
DEFINITION	EST609542 GLSD Medicago truncatula cDNA clone pGLSD-33N16, mRNA
ACCESSION	BQ123966
VERSION	BQ123966.1
KEYWORDS	GI:20175928
SOURCE	EST:
ORGANISM	Medicago truncatula (barrel medic)
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 737)
AUTHORS	Grusak, M.A., Samac, D., Town, C.D., Van Aken, S., Utterback, T., Cheung, F. and Fraser, C.M.
TITLE	Esrs from late stage developing seeds of Medicago truncatula
JOURNAL	Unpublished (2002)
COMMENT	Contact: Grusak, M.A. USDA/ARS Children's Nutrition Research Center


```

REFERENCE
AUTHORS      Zierold,U. and Schweizer,P.
TITLE        Barley ESTs from pathogen-attacked leaf epidermis
JOURNAL      Unpublished (2003)
COMMENT      Contact: Patrick Schweizer
              Transcriptome Analysis, Cytogenetics Department
              Institute of Plant Genetics and Crop Plant Research (IPK)
              Corrensstr. 3, D-06466 Gatersleben, Germany
              Tel: 0049 (0)39482-5660
              Fax: 0049 (0)39482-5595
              Email: schweiz@ipk-gatersleben.de
              Insert Length: 621 Std Error: 0.00
              Plate: 15 row: K column: 4
              Seq primer: SK.
FEATURES
source
1..621
   /organism="Hordeum vulgare"
   /mol_type="mRNA"
   /cultivar="Ingrid BC mlo-5"
   /db_xref="GABI:708556"
   /db_xref="X110-Gold"
   /clone="H015K04"
   /tissue_type="leaf epidermis, 6 h and 24 h post
   inoculation with Blumeria graminis"
   /dev_stage="7 d after germination"
   /lab_host="X110-Gold"
   /clone_lib="HC"
   /note="vector: pbluescript SK+, Site 1: EcoRI (5'-end of
   cDNA); Site 2: XhoI (3'-end of cDNA); Approximately 5 % of
   the clones correspond to cDNA from the fungi B. graminis
   hordei and tritici, respectively. Due to a cloning
   artefact caused by the kit, in most cases the EcoRI site
   is NOT present, as well as the EcoRI adapter used for
   cloning. To excise the insert, restriction sites upstream
   EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
   due to the cloning system used Blue/white selection for
   recombinants is not 100% reliable. Average insert size is
   1.2 kb"
ORIGIN
Alignment Scores:
Pred. No.:      6,79e-103      Length:      621
Score:          985.00      Matches:      190
Percent Similarity: 95.5%      Conservative: 4
Best Local Similarity: 93.60%      Mismatches: 9
Query Match:     66.29%      Indels:      0
DB:              14          Gaps:         0
US-10-723-061-18 (1-292) x CD057603 (1-621)
Qy 46 GlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGlyIleLysSerTyrGlu 65
Db 12 CAGACGTACGTGCGGAACAAGAGCGGTGCGAGCGGTGCGTATCAAGTCGTACGAG 71
Qy 66 ValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsn 85
Db 72 GTCAATTTGCGGAGGACGCTCCGAGTACGAGTCAACAGGCATTCGACCTTCAAT 131
Qy 86 SerAspProSerValHisGlyIleLeuValGlnLeuProLeuProHisMetAsnAsp 105
Db 132 GATGACCCGCTGTGCCATGGCATCTTGGTTGAGTTGCCCTTGCCTGCCATATGACGAC 191
Qy 106 GluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsn 125
Db 192 GAGAACATATTGAATGCTGTAGTATTGAAAGGACGTCGATGGCTTTCATCCAGTGAAT 251
Qy 126 IleGlyArgGluAlaMetGlnGlyArgAspProPhePheValProCysThrProLysGly 145
Db 252 ATTGACCACTGGCGCATCGAGTCCGATCCATCTCTTTGTCATCGACCCCTAAGGA 311
Qy 146 CysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArgAlaValIle 165

```

```

Db 312 TGCATGGAGCTGCTACACAGATCTGGAGTTGAATAATAATGGAAGAGAGCTGTTGTAATT 371
Qy 166 GlyArgSerAsnIleValGlyMetProAlaIleAlaLeuLeuGlnLysAlaAsnAlaThr 185
Db 372 GGAAGAAGCAATATTGTGGGACGCTCTGCTTACTCTTTACAAAAAGCAATGCAACT 431
Qy 186 ValSerIleValHisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIle 205
Db 432 GTGACATTTGACATTTCAAAACCCAGAACCCCGAGGAATAACACAGACAGCATATT 491
Qy 206 ValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTyrIleLysProGlyAla 225
Db 492 ATCATTTGGCGCTGTTGGAGTTGCTAACCTGTTCAGAGGAGTTGATTAAGCCTGAGCT 551
Qy 226 AlaIleIleAspValGlyLysLeuProValAspProGluSerProArgGlyTyrArg 245
Db 552 GCTATTATTGATGTTGGCATCAATCCGTTGATGATCCAGCAAGCCCTCGAGTTATCGT 611
Qy 246 LeuValGly 248
Db 612 CTAGTTGGA 620
RESULT 6
BQ804363
LOCUS      BQ804363
DEFINITION      BQ804363 F06 L11ZS Wheat developing grains cDNA library Triticum
aestivum cDNA clone WHE3553_F06_L11, mRNA sequence.
ACCESSION      BQ804363
VERSION        BQ804363.1
KEYWORDS       832 bp mRNA linear EST 31-JUL-2002
SOURCE         Triticum aestivum (bread wheat)
ORGANISM       Triticum aestivum
REFERENCE       1 (bases 1 to 832)
AUTHORS         Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J.,
Cronin,K., Crossman,C., Fenton,R.D., Iazo,G.R., Pham,J.,
Rausch,C.J., Wilson,C. and Woo,J.
TITLE          The structure and function of the expressed portion of the wheat
              genomes - Developing grains cDNA library
JOURNAL         Unpublished (2002)
COMMENT        Contact: Olin Anderson
              US Department of Agriculture, Agriculture Research Service, Pacific
              West Area, Western Regional Research Center
              800 Buchanan Street, Albany, CA 94710, USA
              Tel: 5105595973
              Fax: 5105595818
              Email: oanderson@w.usda.gov
              Sequences have been trimmed to remove vector sequence and low
              quality sequence with phred score less than 20
              Seq primer: SK primer.
FEATURES
source
1..832
   /organism="Triticum aestivum"
   /mol_type="mRNA"
   /cultivar="Butte 86"
   /db_xref="taxon:4565"
   /clone="WHE3553_F06_L11"
   /tissue_type="Whole grains"
   /dev_stage="3-44 days post anthesis seed"
   /lab_host="E. coli SOLR"
   /clone_lib="Wheat developing grains cDNA library"
   /note="Vector: Lambda ZAP II, excised phagemid: Site 1:
   EcoRI; Plants were grown under six following different
   environmental regimes in greenhouse, Environment 1)
   24oc/17oc day/night, well-watered, with post-anthesis
   fertilizer, Environment 2) 24oc/17oc day/night,
   well-watered, without post-anthesis fertilizer,
   Environment 3) 37oc/17oc day/night, well-watered, with
   post-anthesis fertilizer, Environment 4) 37oc/17oc
   day/night, well-watered, without post-anthesis fertilizer,
   Environment 5) 37oc/17oc day/night plus drought, with

```

post-anthesis fertilizer, Environment 6) 370C/170C day/night plus drought, without post-anthesis fertilizer, and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

ORIGIN

Alignment Scores:
 Pred. No.: 2,33e-102 Length: 832
 Score: 982.00 Matches: 186
 Percent Similarity: 82.90% Conservative: 37
 Best Local Similarity: 69.14% Mismatches: 45
 Query Match: 66.08% Indels: 1
 DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x BQ804363 (1-832)

QY 20 GluileAlaLysMetThyAspAlaLeuGlyValProGlyLeuAlaValLeuVal 39
 Db 5 GAGTCCGCTGCTCTCGTCCGCCACACATCGTCCGGGCTCGGTGTGTATCGT 64
 QY 40 GlySerArgLysAspSerGlnThrTyrValArgAsnLysLysLysLysLysLys 59
 Db 65 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 124
 QY 60 GlyLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLys 79
 Db 125 GGCATCGCTCTTCGACGTGACCTCCCGAGGACATCTCGAGCGCGCTCGTCCG 184
 QY 80 HisLeuAlaThrPheAsnSerAspProSerValHisGlyLeuValGlnLeuProLeu 99
 Db 185 GAGGTCCACCGCTCAACCGCCGACCGCGCCGCGCCGCGCGCGCGCGCGCGCG 244
 QY 100 ProHisHisMetAsnAspGluAsnLeuLeuAsnAlaValSerIleGluLysAspValAsp 119
 Db 245 CCCAGCATATCAGCAGGAATATCTTAACCATGATCTCCATTGAGAGATGTCGAC 304
 QY 120 GlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheVal 139
 Db 305 GGCCTTCATCCTTTGAACATTCGACGTCGCAATGAAAGGAGAGATCCACTTTCGTA 364
 QY 140 ProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLeuLysGly 159
 Db 365 CTTTCACGCCCAAGGATGATGAGGCTCTGTACAGAGTGGCGTCACTGTAAAGGA 424
 QY 160 LysArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaLeuLeuLeu 179
 Db 425 AAACACGCGAGTGTGTGGGGTGAACATCATCGTGGGTTTACCATGATCTCTCTCTG 484
 QY 180 GlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluLeu 199
 Db 485 CTGAAGCGGAGCGGACCGCTGTCATGTCGATTCATTCAGGACCCCAATTCGGAACAT 544
 QY 200 ThrArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgLysSer 219
 Db 545 GTCCGTCAGCAGACATTTGTCATTGACAGCGTGGCCGCGCATGATCAAGGAGAC 604
 QY 220 TrpIleLysProGlyAlaAlaIleLeuAspValGlyIleAsnProValAspAspProGlu 239
 Db 605 TGGATCAACACGCGCGGCTGATAGCTGGGCAAACTTCATCGACGACCCCAACC 664

QY 240 SerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAla 259
 Db 665 AGGAAGCTGGGTACACACTCGTTGGCGATGTGGATTTCTCGGAGCGAGCAAGGTGCG 724
 QY 260 GlyLeuIleThrProValProGlyGlyValGlyProMet-ThrIleAlaMetLeuLeu 279
 Db 725 GGTCACTTGACTCGGTCCCGGAGGCGTGGGCCGATGAACCGTGGCGATGTGTGTGA 784
 QY 279 rAsnThrLeuGluSexAlaLysArg 287
 Db 785 GAACACGCTGCGACGGCGCCAGAGG 809

RESULT 7

CD574828/c

LOCUS

DEFINITION

AGI Poncirus trifoliata cDNA clone UCRPT01_01dd09, mRNA sequence.

ACCESSION

CD574828

KEYWORDS

SOURCE

ORGANISM

Poncirus trifoliata

Poncirus trifoliata

Spematophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

1 (bases 1 to 838)

REFERENCE

AUTHORS

Wanamaker, S., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,

Wing, R. and Yu, Y.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata CTV-challenged phloem - AGI

Unpublished (2003)

Contact: Mikeal Roose

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124, USA

Tel: 9097874137

Fax: 9097874437

Email: mikeal.roose@ucr.edu

Seq primer: T7.

Location/Qualifiers

1..838

/organism="Poncirus trifoliata"

/mol_type="mRNA"

/cultivar="Pomeroy Op"

/db_xref="taxon:37690"

/clone="UCRPT01_01dd09"

/tissue_type="Phloem"

/dev_stage="10 - 30 cm shoots"

/lab_host="E. coli TUC121"

/clone_lib="Poncirus trifoliata CTV-challenged cDNA

library - AGI"

/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: BcoRI; Site 2: XhoI; plants were grown in the

greenhouse at University of California, Riverside. The

seeds were open-pollinated (very probably selfed)

seedling of Poncirus trifoliata cv Pomeroy that was

selected as homozygous for the CTV resistance gene. The

rootstock was sweet orange infected with citrus tristeza

virus (CTV) isolate T514 over 1 year before sampling (CTV

infests sweet orange, but not genotypes carrying the CTV

resistance gene. Shoots 10-30 cm long were harvested in

October 2003, and the green phloem (bark) was removed and

frozen quickly in dry ice. Total RNA was extracted using

Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA

library was made, and 0.5 million primary lambda cDNA

clones were in vivo excised to give a population of

phagescript SK(-) phagemids. All steps to this point were

performed in the ML Roose lab at the University of

California, Riverside by X. Ye. Phagemids were plated,

plasmid DNA purified, cDNA clones archived, and DNA

sequences determined bi-directionally using an ABI3730 at

the Arizona Genomics Institute, University of Arizona

(Collura, Feuerbacher, Kim, Kudrna, Wing, Yu).
Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanmaker) using the HarVest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

ORIGIN

Alignment Scores:
Pred. No.: 2.35e-102 Length: 838
Score: 982.00 Matches: 183
Percent Similarity: 92.04% Conservative: 25
Best Local Similarity: 80.97% Mismatches: 18
Query Match: 66.08% Indels: 0
DB: 14 Gaps: 0

US-10-723-061-18 (1-292) x CD574828 (1-838)

Qy 67 AsnLeuProGluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSer 86
Db 837 CATTTACTGAGGACCTCTGAAACAAGAGTACTCAAGATATTTTCAGTCTTCATGAT 778
Qy 87 AspProSerValHisGlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGlu 106
Db 777 GATCCTTCAGTTCATGGCATCCTTGTGCAGTTGCTCTGCTCTCATATAGATGAGCAG 718
Qy 107 AsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsnIle 126
Db 717 AGCATCTTAAATGCTGCAGATGAGAAAGATGTCAGCGTTTCCACCGCTGAATATT 658
Qy 127 GlyArgLeuAlaMetGlnGlyArgAspProPhePheValProCysThrProLysGlyCys 146
Db 657 GGCGCTGAGCATGCGAGGTAGAGACCTTGTTCATCTTGTACGCTTAAGGTTGC 598
Qy 147 MetGluLeuLeuHisArgTyGlyValGluLysGlyLysArgAlaValValIleGly 166
Db 597 ATAGAGTTGTTCATAGATATGGCTTTCATATCAAGGAAAGAGGCTGTCTGATTTGGC 538
Qy 167 ArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLysAlaAsnAlaThrVal 186
Db 537 CGAAGCATATTTGTGGATGCCAGCTGTTTGTCTCCAAAGGGAAGTCTACTGTT 478
Qy 187 SerIleValHisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIleVal 206
Db 477 AGTATTTCCATTCAGAACTAAGAACCTGAGGAGATTACGACAGCGCATATCATA 418
Qy 207 IleAlaAlaValGlyValAlaAsnLeuValArgGlySerThrIleLysProGlyAlaAla 226
Db 417 ATTTCCGCTGTAGGCAACCAATATGTTAAGGGGGAGCTGGATAAACCTGGTGCACTC 358
Qy 227 IleIleAspValGlyIleAsnProValAspAspProGluSerProArgGlyTyArgLeu 246
Db 357 ATAATTGATTTGGATTAACCCACTCGAGGATGCAAAAGTCCCGAGGTTACCGTTG 298
Qy 247 ValGlyAspValCysTyrcGluAlaSerIleAlaGlyLeuIleThrProValPro 266
Db 297 GTTGAGATGTTTGTACAGAGGCATGTGAGGTTGCTTCAGCTATACATCCAGTTCCT 238
Qy 267 GlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGluSerAlaLys 286
Db 237 GGAGAGTTTGGTCCATGACTATAGCATGTCTCTCTTAATACCTCACTTCGCGAAG 178
Qy 287 ArgIleHisLysPheLys 292
Db 177 AGGGTGCAACAATTCCAG 160

RESULT 8

BE359332 605 bp mRNA linear EST 20-JUL-2000
LOCUS DG1_40.F10.b1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
DEFINITION sequence.

ACCESSION BE359332
VERSION BE359332.1 GI:9300889
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1. (bases 1 to 605)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and
Pratt,L.H.

AUTHORS

An EST database from Sorghum: dark-grown seedlings

TITLE

Unpublished (2000)

JOURNAL

Contact: Cordonnier-Pratt MM

COMMENT

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence

is 20.

Seq primer: JEN REV

High quality sequence stop: 599

POLYA=No.

Location/Qualifiers

1..605

/organism="Sorghum bicolor"

/mol_type="mRNA"

/db_xref="taxon:4558"

/clone_lib="Dark Grown 1 (DG1)"

/note="Organ: 5-day-old dark-grown seedlings; Vector:
Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Clones to be sequenced were prepared by mass excision."

Db 369 AACCTGGAGCAGTATTATGATGTTGGTATCAACCCAGTTCATGCCAGAAACCCCT 428
 QY 242 ArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGlyLeu 261
 Db 429 CGGGTTTATAGGCTGTCGAGAGTGTGTTATGAGGAGCCCTCAAGGTGCGGAGCA 488
 QY 262 IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuSerAsnThr 281
 Db 489 ATCACACAGTTCAGAGGTGTGGCCCATGACAAATGCAATGCTGTTTGTCAACACA 548
 QY 282 LeuGluSerAlaLysArgIleHisLysPheLys 292
 Db 549 CTTGAGTCACTAAAGAAATCCAAATTCAA 581

RESULT 9
 BG415373
 LOCUS
 DEFINITION
 HVSMEK0006D16f Hordeum vulgare testa/pericarp EST library
 HVCNDA0013 (normal) Hordeum vulgare subsp. vulgare cDNA clone
 HVSMEK0006D16f, mRNA sequence.
 BG415373
 BG415373.1 GI:13320924
 EST.
 Hordeum vulgare subsp. vulgare
 Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Scrotophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poodeae; Triticeae; Hordeum.
 1 (bases 1 to 533)
 Wing, R., Close, T.J., Klein, H., Wise, R., Kannangara, G., von
 Wetstein, D., Akhunov, E., Chin, A., Choi, D.W., Fenton, R.D.,
 Kianian, P., Otto, C., Simons, K., Zhang, D., Begum, D., Frisch, D.,
 Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and
 Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex testa/pericarp cDNA library
 Unpublished (2001)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Total hg bases = 267
 Seq primer: AATTAAACCTCACTAAAGGG
 High quality sequence start: 8
 High quality sequence stop: 598.
 Location/Qualifiers
 1..633
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSMEK0006D16f"
 /tissue_type="testa/pericarp"
 /lab_host="TUC121"
 /clone_lib="Hordeum vulgare testa/pericarp EST library
 HVCNDA0013 (normal)"

/note="Vector: lambdaZAP; Site 1: EcoR1; Site 2: Xho1;
 plants were raised from seeds in a Controlled Environments
 growth chamber maintained in continuous light at 18°C, and
 testa and pericarp were dissected from developing kernels
 at Washington State University, Pullman, WA (Kannangara,
 von Wetstein). Total RNA was prepared, poly(A) RNA was
 purified, one cDNA library was made, and 1 million pfu
 were in vivo excised to give plasmidscript SK(-) cDNA
 phagemids in the TJ Close lab at the University of
 California, Riverside (Akhunov, Chin, Choi, Close, Fenton,
 Kianian, Otto, Simons, Zhang). Phagemids were plated and
 picked at the Clemson University Genomics Institute (CUGI)

(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
 preparations, DNA sequencing and sequence analysis were
 performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
 Oates, Rambo, Main). The sequence has been trimmed to
 remove vector sequence and contains a minimum of 100 bases
 of phred value 20 or above. For more details on library
 preparation and sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders/Also
 see Close TJ, Wing R, Klein, H., Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/ggpages/bgr/31/cover.html)"

ORIGIN

Alignment Scores:
 Pred. No.: 2,89e-98 Length: 633
 Score: 945.00 Matches: 186
 Percent Similarity: 91.39% Conservative: 5
 Best Local Similarity: 89.00% Mismatches: 18
 Query Match: 63.59% Indels: 0
 DB: 12 Gaps: 0
 US-10-723-061-18 (1-292) x BG415373 (1-633)
 QY 30 ValValProGlyLeuAlaValIleLeuValGlySerArgLysAspSerGlnThrTyrVal 49
 Db 6 GTGGTGCTGAGTGGCGGCCATCTAGTCGGCTCGAGGAACGATCCCATCTACTACGTG 65
 QY 50 ArgAsnLysLysLysAlaCysGluAlaValGlyLysSerTyrGluValAsnLeuPro 69
 Db 66 CGGAACAGAGAGAGCGGCGGCGAGCGGTCTGATCAAGTCGTACAGAGTCATTTGCCG 125
 QY 70 GluAspSerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSer 89
 Db 126 GAGGAGCGTTCGAGTACTAGGTCTCAAGACATTCGACCTTCAATGATGACCGTCT 185
 QY 90 ValHisGlyLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnLeu 109
 Db 186 GTTCATGGCATCTGGGTTCATTTGCCCTTGCCTCGCCATATGAACGACGACATATTG 245
 QY 110 AsnAlaValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeu 129
 Db 246 AATGCTGTAGTATTGAAAAGGAGCGTCGATGGCTTCATCCAGTGAATATTGGACACTG 305
 QY 130 AlaMetGlnGlyArgAspProPheValProCysThrProLysGlyCysMetGluLeu 149
 Db 306 GCGATGCAAGGTGCGGATCCATCTTTGTTCATGCCCTTAAAGGATGCGATGAGCTG 365
 QY 150 LeuHisArgTyrGlyValGluLysGlyLysAlaValValIleGlyArgSerAsn 169
 Db 366 CTACACAGATCTGGAGTTCGAATAAATGGAAGAGAGCTGTTGTAATGGAGAAGCAAT 425
 QY 170 IleValGlyMetProAlaIleLeuLeuGlnLysAlaAsnAlaThrValSerIleVal 189
 Db 426 ATTGTGGGACCGCTGCTCTTACTCTTACAAAAGCAAAATGCAACTGTGAGCATTTGTA 485
 QY 190 HisSerAsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAla 209
 Db 486 CATTCAAAACCAAAACCCCGAGGAATAACAACAGCAGAGATTTTATCATTCGGCT 545
 QY 210 ValGlyValAlaAsnLeuValArgGlySerThrIleLysProGlyAlaAlaIleLeu 229
 Db 546 GTTGGAGTTGCTAACTGGTCTAGAGGAGTTGGATAAAGCCCTGGAGCTGCTATTATCAT 605
 QY 230 ValGlyIleAsnProValAspAspPro 238
 Db 606 GTTGGCATCAATCCCGTTGATGATCCA 632

RESULT 10

CK298478/c

LOCUS

DEFINITION

CK298478 870 bp mRNA linear EST 15-DEC-2003
 EST761192 Nicotiana benthamiana mixed tissue cDNA library,

normalized, full-length Nicotiana benthamiana cDNA clone NEMDG17 3' end, mRNA sequence.

ACCESSION CK298478.1 GI:39885892

VERSION EST.

KEYWORDS Nicotiana benthamiana

SOURCE Nicotiana benthamiana

ORGANISM Nicotiana benthamiana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

TITLE asterids; lamids; Solanales; Solanaceae; Nicotiana.

JOURNAL 1 (bases 1 to 870)

COMMENT Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A., Day, B.,

Skaskawicz, B., Jin, H. and Baker, B.

Generation of EST sequences from Nicotiana benthamiana

Other ESTs: EST761191

Contact: Robin Buell

The Institute for Genomic Research

9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potato-array@tigr.org

Clones can be requested from TIGR via potato@tigr.org

Seq primer: GTA ATA CGA CTC ACT ATA GGG C.

Location/Qualifiers

1..870

/organism="Nicotiana benthamiana"

/mol_type="mRNA"

/db_xref="taxon:4100"

/clone="NEMDG17"

/tissue_type="abiotic and biotic stress-treated leaves,

callus tissue and root tissue"

/lab_host="DH10B-Tona"

/clone_lib="Nicotiana benthamiana mixed tissue cDNA

library, normalized, full-length"

/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;

supplier: RNA was isolated from Nicotiana benthamiana

tissues that include callus, roots from liquid culture

grown plants, heat-stressed leaves (38 C, 3 hr and 6 hr),

cold-stressed leaves (5 C 3 hr, 6hr), and pathogen

challenged leaves (Pseudomonas syringae pv tomato 12 hr;

Xanthomonas campestris pv campestris 12 hr, 18hr;

Pseudomonas syringae pv phaseolicola 18hr, and Xanthomonas

campestris pv vesicatoria 18hr). RNA was isolated from

these tissues and pooled in approximately equal molar

amounts."

ORIGIN

Alignment Scores:

Pred. No.: 8, 6e-97 Length: 870

Score: 934.00 Matches: 178

Percent Similarity: 90.65% Conservative: 16

Best Local Similarity: 83.18% Mismatches: 20

Query Match: 62.85% Indels: 0

DB: 14 Gaps: 0

US-10-723-061-18 (1-292) x CK298478 (1-870)

QY 78 LeuLysHisIleAlaThrPheAsnSerAspProSerValHisGlyLeuValGlnLeu 97

DB 870 CTCAGTATATCTCAGGCTTCAATGATGATCTTATGTTTCATGATTTCTGTTTCAGCTA 811

QY 98 ProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAsp 117

DB 810 CCGTTACCTTCTCATATGAATGAACAGAACATCTTAAATGCTGTTTGCATTGAGAAGGAT 751

QY 118 ValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhe 137

DB 750 GTGGATGGATTCCACCCTAAATATTCGCCGCTAGCGATCGGAGTAGAAGCACTA 691

QY 138 PheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIle 157

DB 690 TTGTGTCATGTACACCTTAAGGATCGATTGAGCTACTGCTCGGTATTAATGTGGAGATC 631

QY 158 LysGlyLysArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaIleLeu 177

DB 630 AAGCGGAAAGAGCTGTGTGTTGGCAGGAGTATATATCGTCGGAACCGCGCTCTCTG 571

QY 178 LeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGlu 197

DB 570 CTTCTGCAGAGGGAAGATGCCACTGTGACGATTGTCATTCAGAACCAAGAACTGAG 511

QY 198 GluIleThrArgGlnAlaAspIleValIleAlaValAlaGlyValAlaAsnLeuValArg 217

DB 510 GAGATCAGCAGACAGCAGATATCATATCTTCAGCTGTAGTCAACCAACATGTTTGA 451

QY 218 GlySerTrpIleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAsp 237

DB 450 GGCAGCTGGATCAAGCTGGGCTGTGATTATGATGTCGGGATCAACCTGTTGAGAT 391

QY 238 ProGluSerProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLys 257

DB 390 GCTACAAATCTCGCGCTATCGCTAGTTGGAGATGTTTGTATGAGAGGCCAGCAA 331

QY 258 IleAlaGlyLeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeu 277

DB 330 GTTGCTTCAGCTATCACTCTCTCTCTGGGGAGTTGGACCATGACTATAGCAATGCTT 271

QY 278 LeuSerAsnThrLeuGluSerAlaLysArgIleHisLysPhe 291

DB 270 CTTTCTATATCTTGTATCAGAAAGCGAATTCACACTTC 229

CA008897 586 bp mRNA linear EST 23-OCT-2002

HUI2H11r HU Hordeum vulgare subsp. vulgare cDNA clone HUI2H11

5-PRIME, mRNA sequence.

CA008897.1 GI:24285879

EST.

Hordeum vulgare subsp. vulgare

Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Hordeum.

1 (bases 1 to 586)

Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and

Graser, A.

Barley ESTs from germinating seeds

Unpublished (2002)

Contact: Stein Nils

Molecular Markers Group, Department Genbank

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@ipk-gatersleben.de

Insert length: 586 Std Error: 0.00

Plate: 12 row: H column: 11

Seq primer: M13rev.

Location/Qualifiers

1..586

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="barke"

/sub_species="vulgare"

/db_xref="GABI:266547"

/db_xref="taxon:112509"

/clone="HUI2H11"

/tissue_type="germinating seeds"

/dev_stage="germinating seeds, 16-48 h"

/lab_host="XL10-Gold"

/clone_lib="HU"

/notes="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of

cDNA); Site 2: XhoI (3'-end of cDNA); germinating seeds,

16-48h. Due to a cloning artefact caused by the kit, in

most cases the EcoRI site is NOT present, as well as the

EcoRI adapter used for cloning. To excise the insert,

restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"

ORIGIN

Alignment Scores:
 Pred. No.: 5,23e-96 Length: 586
 Score: 925.00 Matches: 179
 Percent Similarity: 94.87% Conservative: 6
 Best Local Similarity: 91.79% Mismatches: 10
 Query Match: 62.25% Indels: 0
 DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x CA008897 (1-586)

Qy 42 ArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGlyIle 61
 Db 1 CGGCAGAGCCAGAGCTACGTCCGGACACAGAGAGCGGTGCGAGCGGTGATC 60
 Qy 62 LysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHisIle 81
 Db 61 AAGTCGTACGAGTCAATTTCGCGAGGAGCGCTCCGAGTACGAGTCAATCAAGCACAT 120
 Qy 82 AlaThrPheAsnSerAspProSerValHisGlyLeuValGlnLeuProLeuProHis 101
 Db 121 GCACCTTCATGATGACCGCTGTCCATGGCATCTTGGTTCAGTTCGCCCTTGCTCGC 180
 Qy 102 HisMetAsnAspGluAsnLeuLeuAsnAlaValSerIleGluLysAspValAspGlyPhe 121
 Db 181 CATATCAACGACGAGAACATATTGAATGCTGTATTGTAAGAGGAGCGTGCATT 240
 Qy 122 HisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCys 141
 Db 241 CATCCAGTGAATATGGAGACTGGCGATGCGAGTGCAGAGTCCGATTCCTTTCCATGC 300
 Qy 142 ThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLysArg 161
 Db 301 ACCCTAAAGGATGATGAGCTGCTACACAGATCTGGAGTTGAATTAATGGAGAGA 360
 Qy 162 AlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGlnLys 181
 Db 361 GCTGTGTAAATGGAGAGCAATATTGTGGGAGCGCTGCTGCTTACTCTTACAAAA 420
 Qy 182 AlaAsnAlaThrValSerIleValHisSerAsnThrLysProGluGluIleThrArg 201
 Db 421 GCATATGCACTGTGAGCATTTGTACATTCAAAAACAGAACCCCGAGAGAAATACAGA 480
 Qy 202 GlnAlaAspIleValIleAlaValGlyValAlaAlaAsnLeuValArgIleSerTrpIle 221
 Db 481 CAAGCAGATATTATTTCGCGCTGTGGAGTTGTCTAACCTGGTCAGAGGAGTTGGATA 540
 Qy 222 LysProGlyAlaAlaIleIleAspValGlyIleAsnProValAsp 236
 Db 541 AAGCTGAGCTGCTATTATTGATTTGTCATCAATCCGGTTGAT 585

RESULT 12

BQ609202 595 bp mRNA linear EST 25-JUN-2002
 LOCUS BRY 5126 wheat EST endosperm library Triticum aestivum cDNA 5',
 mRNA sequence.

ACCESSION BQ609202
 VERSION BQ609202.1 GI:21558541
 KEYWORDS EST.

SOURCE

Triticum aestivum (bread wheat)
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.

REFERENCE

1 (bases 1 to 595)
 Clarke,B., Lambrecht,M. and Rhee,S.Y.
 Arabidopsis genomic information for interpreting wheat EST
 sequences

JOURNAL

MEDLINE

PUBMED

COMMENT

Funct. Integr. Genomics 3 (1-2), 33-38 (2003)

22478026

12590341

Contact: Lambrecht M

The Arabidopsis Information Resource

Carnegie Institution of Washington, Dept. of Plant Biology

260 Panama Street, Stanford, CA 94305, USA

Tel: 1 650 325 1521 x 251

Fax: 1 650 325 3748

Email: rheesacoma.stanford.edu.

Location/Qualifiers

1..595

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="Wvuna"

/db_xref="taxon:4565"

/tissue_types="endosperm"

/dev_stage="developing endosperm tissue 8, 10 and 12 DPA

(days post anthesis)"

/clone_lib="wheat EST endosperm library"

ORIGIN

Alignment Scores:
 Pred. No.: 6.98e-96 Length: 595
 Score: 924.00 Matches: 182
 Percent Similarity: 93.97% Conservative: 5
 Best Local Similarity: 91.46% Mismatches: 10
 Query Match: 62.18% Indels: 2
 DB: 13 Gaps: 0

US-10-723-061-18 (1-292) x BQ609202 (1-595)

Qy 72 SerSerGluAspGluValLeuLysHisIleAlaThrPheAsnSerAspProSerValHis 91
 Db 3 GCTTCCGAGTACGAGGTCAATCAAGCACATAGCAACCTTCAACGATCACCGTCGGTGCAT 62
 Qy 92 GlyIleLeuValGlnLeuProLeuProHisHisMetAsnAspGluAsnIleLeuAsnAla 111
 Db 63 GGCATCTCGTTTCAGCTGCCCTTACCCTCACCATATCAACGACGAGAACATATTGAATGCT 122
 Qy 112 ValSerIleGluLysAspValAspGlyPheHisProLeuAsnIleGlyArgLeuAlaMet 131
 Db 123 GTTAGTATTGAAAGGATGTTGATGCTTTCATCCAGTGAACATTCGACGACTGCGATG 182
 Qy 132 GlnGlyArgAspProPhePheValProCysThrProLysGlyCysMetGluLeuHis 151
 Db 183 CAAGGTGGGATCCATCTTGTTCATGACCCCTAAAGAGTGCATGGAGCTGTACAC 242
 Qy 152 ArgTyrGlyValGluIleLysGlyLysArgAlaValValIleGlyArgSerAsnIleVal 171
 Db 243 AGATCTGGAGTTGAATTAAGGAGAGAGAGCTGTTGTAATTTGGAAGGACCAATATTGTG 302
 Qy 172 GlyMetProAlaAlaLeuLeuGlnLysAlaAsnAlaThrValSerIleValHisSer 191
 Db 303 GGGACCGCTGCTGCTTACTCTTACAAAAAGCAAAATGCAACTGTGAGTATTGTACATTCA 362
 Qy 192 AsnThrLysLysProGluGluIleThrArgGlnAlaAspIleValIleAlaValGly 211
 Db 363 AAACCCAGAACCCCGAGGAATATACAGACAGAGCATATTATATCCGGCTGTGGA 422
 Qy 212 ValAlaAsnLeuValArgGlySerTrpIleLysProGlyAlaAlaIleIleAspValGly 231
 Db 423 GTTGTAACTCGTTCAGAGGAGTTGGATAAAGCCCTGGAGCTGCTATTATTGATGTGCGC 482
 Qy 232 IleAsnProValAspAspProGluSerProArgGlyTyrArgLeuValGlyAspValCys 251
 Db 483 ATCAATCCGGTTGATGATCCAGCAAGTCTCCAGGTTTACCGGCTAGTTGGAGATGCTGCG 542
 Qy 252 TyrGluGluAlaSerLys-IleAlaGlyLeuIleThrPro-ValProGlyGly 268
 Db 543 TTGTAGGAAGCCTNCAAGGNTGCAGGAGCCATCACACAGGTCCCGCGCGG 595

RESULT 13

```
BQ295278
LOCUS       BQ295278               762 bp     mRNA          linear      EST 16-MAY-2002
DEFINITION   Wheat unstressed root tip cDNA library Triticum aestivum
ACCESSION   BQ295278
VERSION     BQ295278.1  GI:20810800
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM    Triticum aestivum
REFERENCE   1 (bases 1 to 762)
AUTHORS     Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C.,
            Gustafson,P., Lazo,G.R., Pham,J., Rausch,C.J., Ross,K., Wilson,C.
            and Woo,J.
TITLE       The structure and function of the expressed portion of the wheat
            genomes - Unstressed root tip cDNA library
JOURNAL     Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             source
     1..762
     /organism="Triticum aestivum"
     /mol_type="mRNA"
     /cultivar="Chinese Spring"
     /db_xref="taxon:4565"
     /clone="WHE2868_B10_C20"
     /tissue_type="Root tip at 1.0 to 1.5 mm stage"
     /dev_stage="Four-day old seedling"
     /lab_host="E. coli SOLR"
     /clone_lib="Wheat unstressed root tip cDNA library"
     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
            Site 1: EcoRI; Site 2: XhoI; Plants were grown under
            hydroponic conditions for four days. Root tips were
            excised and snap frozen (Ross and Gustafson) and total RNA
            was prepared at University of Missouri, Columbia. Poly(A)
            RNA was purified, a cDNA library was made, and the cDNA
            clones were in vivo excised to give phagescript
            SK(-) phagemids in the TU Close lab (Chin and Close) at the
            University of California, Riverside. Plasmid DNA
            preparations and DNA sequencing were performed in the OD
            Anderson lab (all other authors)."
```

```
Alignment Scores:
Pred. No.:      2.44e-94      Length:      762
Score:          912.00      Matches:      178
Percent Similarity: 96.8%      Conservative: 7
Best Local Similarity: 93.1%      Mismatches: 6
Query Match:    61.37%      Indels:      0
DB:             13          Gaps:        0

US-10-723-061-18 (1-292) x BQ295278 (1-762)

QY 1 AlaLysIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
|||||
DB 189 GCARAGATCATCGACGGAGAGTGTGGCGAAGCAATAAGAGAGGAGTCGGCGTTGAG 248
|||||
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
|||||
DB 249 ATCGCCCGGATCAGACGCGATCGCGGTGTGCTGGCGTGGCGGTTCATCTAGTCGCG 308
|||||
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
|||||

309 TCGAGGAAGGATTCCACGACCTACCTGCGGAACAAGAGAGGGCTGTGATCGGTCGGT 368
61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
|||||
369 ATCAAGTCGTATGAGGTCAATCTGCCAGAGGACGCTTCGGAGTACGAGGTCAATCAAGCAC 428
81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
|||||
429 ATAGCAACCTTCAACGATGATCCGTCGGTGCATGCGCATCTTGGTTTCAGTCGCCCTTACT 488
101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
489 CGCCATATGACGACGAGACATATTTGAATGCTGTAGTATTGAGAGGACGCTTGTATGGC 548
121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
|||||
549 TTTCATCCAGTGAACATTGCGACGCTTGCATGCAAGGTTCGGGATCCCATCTTTGTTCAC 608
141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
609 TGACCCCTTAAGGATGCGATGAGCTGCTACACATCTGGAGTTGAAATAAAGGAAG 668
161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
669 AGAGCTGTTGTAATTGGAAGGAGCAATATTGTGGGACACCTGCTGCTTACTCTTACAA 728
181 LysAlaAsnAlaThrValSerIleValHisSer 191
729 AAAGCAATGCAACTGTGATGATTGTACATTCA 761

RESULT 14
BQ788577
LOCUS       BQ788577               708 bp     mRNA          linear      EST 26-JUL-2002
DEFINITION   WHE4151_D10_G19ZS Wheat CS whole plant cDNA library Triticum
ACCESSION   BQ788577
VERSION     BQ788577.1  GI:21997049
KEYWORDS    EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM    Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 708)
AUTHORS     Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K.,
            Dvorak,J., Lazo,G.R., Rausch,C.J., Wilson,C. and Woo,J.
TITLE       The structure and function of the expressed portion of the wheat
            genomes - Chinese Spring whole plant cDNA library
JOURNAL     Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             source
     1..708
     /organism="Triticum aestivum"
     /mol_type="mRNA"
     /cultivar="Chinese Spring"
     /db_xref="taxon:4565"
     /clone="WHE4151_D10_G19"
     /tissue_type="Roots, leaves, crown, stem and sheath"
     /dev_stage="Adult"
     /lab_host="E. coli SOLR"
     /clone_lib="Wheat CS whole plant cDNA library"
     /note="Vector: Lambda Uni-ZAP XR, excised phagemid
            phagescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant
            tissues from wheat cv. CS grown to full tillering stage in
```


[illegible]

Search completed: July 24, 2004, 09:45:06
Job time : 3415 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:39:06 ; Search time 969 Seconds
(without alignments)
94.366 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIIDGKLVAKQIREIAVE.....TIAMLLSNTLESAKRIHKFX 292

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1288442 seqs, 313154207 residues

Total number of hits satisfying chosen parameters: 1288442

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1486	100.0	292	9	US-09-903-814-18
2	1486	100.0	292	16	US-10-723-061-18
3	1234	83.0	342	12	US-10-424-599-234064
4	1030	69.3	242	9	US-09-903-814-20
5	1030	69.3	242	16	US-10-723-061-20
6	1025	69.0	292	12	US-10-425-114-60605
7	1025	69.0	299	12	US-10-425-114-56838
8	1019	68.6	361	12	US-10-424-599-257130
9	1019	68.6	373	12	US-10-425-114-45924
10	1016	68.4	290	9	US-09-903-814-16
11	1016	68.4	290	16	US-10-723-061-16
12	1012	68.1	292	16	US-10-437-963-176965
13	1010	68.0	294	12	US-10-424-599-268535
14	994	66.9	321	12	US-10-425-114-44200
15	988	66.5	395	16	US-10-437-963-125501

ALIGNMENTS

RESULT 1

```
US-09-903-814-18
; Sequence 18, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Ramodu, Layo O.
; APPLICANT: Crozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-903-814-18
```

```
Query Match 100.0%; Score 1486; DB 9; Length 292;
Best Local Similarity 100.0%; Pred No. 1.6e-145;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AKIIDGKLVAKQIREIAVEIAKMDAIGVPGGLAVILGSRKDSQTYVRNKKCAEAVG 60
Db 1 AKIIDGKLVAKQIREIAVEIAKMDAIGVPGGLAVILGSRKDSQTYVRNKKCAEAVG 60
Qy 61 IKSEYVNLPEDESSEDLVLAHATFNSDPVGHILVOLPLPHMNDENILNAVSIKDVVG 120
Db 61 IKSEYVNLPEDESSEDLVLAHATFNSDPVGHILVOLPLPHMNDENILNAVSIKDVVG 120
Qy 121 FHPNLIGRLAMQGRDFFFPVCTPKGCMELHRYGVIEKGRVAVIGRSNIVGPAALLIQ 180
```

```
16 965 64.9 396 12 US-10-425-114-65365 Sequence 65365, A
17 919 61.8 605 16 US-10-437-963-155078 Sequence 155078, A
18 801 53.9 220 12 US-10-425-114-72298 Sequence 72298, A
19 775 52.2 279 15 US-10-369-493-10151 Sequence 10151, A
20 735 49.5 935 16 US-10-408-765A-33 Sequence 33, Appl
21 730 49.1 344 15 US-10-369-493-16864 Sequence 16864, A
22 724 48.7 285 10 US-09-769-787-31 Sequence 31, Appl
23 722.5 48.6 297 15 US-10-369-493-11822 Sequence 21822, A
24 716 48.2 279 15 US-10-369-493-17359 Sequence 17359, A
25 715 48.1 299 15 US-10-369-493-12212 Sequence 12212, A
26 711 47.8 285 15 US-10-369-493-19483 Sequence 19483, A
27 710.5 47.8 310 15 US-10-369-493-19300 Sequence 19300, A
28 710 47.8 286 12 US-10-424-599-268512 Sequence 268512, A
29 707.5 47.6 294 15 US-10-369-493-18390 Sequence 18390, A
30 701 47.2 289 9 US-09-903-814-22 Sequence 22, Appl
31 701 47.2 289 16 US-10-723-061-22 Sequence 22, Appl
32 694.5 46.7 278 15 US-10-369-493-17921 Sequence 17921, A
33 690.5 46.5 299 15 US-10-369-493-8083 Sequence 8083, Ap
34 687.5 46.3 284 15 US-10-369-493-10872 Sequence 10872, A
35 684.5 46.1 280 15 US-10-369-493-16550 Sequence 16550, A
36 683.5 46.0 280 15 US-10-369-493-7153 Sequence 7153, Ap
37 683.5 46.0 283 15 US-10-369-493-23202 Sequence 23202, A
38 681 45.8 285 15 US-10-369-493-10348 Sequence 10348, A
39 679 45.7 288 15 US-10-369-493-744 Sequence 744, Ap
40 678.5 45.7 279 15 US-10-369-493-4398 Sequence 4398, Ap
41 675 45.4 282 15 US-10-369-493-19921 Sequence 19921, A
42 673.5 45.3 286 15 US-10-369-493-20586 Sequence 20586, A
43 670 45.1 283 15 US-10-369-493-14053 Sequence 14053, A
44 668 45.0 284 15 US-10-369-493-18751 Sequence 18751, A
45 668 45.0 291 15 US-10-369-493-146 Sequence 146, App
```

Db 121 FHLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIVGMPAALLQ 180
QY 181 KANATSVIHSNTKPEEITRQADIVIAAVGVANLVGRSWIKRPGAAIDVGINPVDPPES 240
Db 181 KANATSVIHSNTKPEEITRQADIVIAAVGVANLVGRSWIKRPGAAIDVGINPVDPPES 240
QY 241 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 292
Db 241 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 292

RESULT 2
US-10-723-061-18
; Sequence 18, Application US/10723061
; Publication No. US20040132150A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV1
; CURRENT APPLICATION NUMBER: US/10/723,061
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: US 09/903,814
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-723-061-18

Query Match 100.0%; Score 1486; DB 16; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.6e-145;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAQIREIAVEIAAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKACEAVG 60
Db 1 AKIIDGKLVAQIREIAVEIAAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKACEAVG 60
QY 61 IKSEYVNLPEDESSEDEVLKHIAFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVGD 120
Db 61 IKSEYVNLPEDESSEDEVLKHIAFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVGD 120
QY 121 FHLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIVGMPAALLQ 180
Db 121 FHLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIVGMPAALLQ 180
QY 181 KANATSVIHSNTKPEEITRQADIVIAAVGVANLVGRSWIKRPGAAIDVGINPVDPPES 240
Db 181 KANATSVIHSNTKPEEITRQADIVIAAVGVANLVGRSWIKRPGAAIDVGINPVDPPES 240
QY 241 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 292
Db 241 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 292

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234064
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53384C.1.pep
US-10-424-599-234064

Query Match 83.0%; Score 1234; DB 12; Length 342;
Best Local Similarity 79.8%; Pred. No. 3e-119;
Matches 233; Conservative 29; Mismatches 30; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAQIREIAVEIAAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKACEAVG 60
Db 51 AKVIDGKSAQIREDeltaEVSRRRESIGVPGGLAVILVGRSKDSQTYVRNKKKACEAVG 110
QY 61 IKSEYVNLPEDESSEDEVLKHIAFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVGD 120
Db 111 INSLEANLPEDSTEESEVLYAGYNDPSVHGILVOLPLPHMNEQNILNAVRIEKDVGD 170
QY 121 FHLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIVGMPAALLQ 180
Db 171 FHLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIVGMPAALLQ 180
QY 181 KANATSVIHSNTKPEEITRQADIVIAAVGVANLVGRSWIKRPGAAIDVGINPVDPPES 240
Db 231 REDATSVIHSRTSNPEEITRQADIIIAAGVQANVGRSWIKRPGAVIIDVGINPVDPPES 290
QY 241 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 292
Db 291 PRGYRLVGDVVCYEASKEIAGLITPVGCGVGPMTIAMLSNTLESKRHHKFK 342

RESULT 4
US-09-903-814-20
; Sequence 20, Application US/09903814
; Patent No. US20020102689A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB-1179-A
; CURRENT APPLICATION NUMBER: US/09/903,814
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/092,869
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Glycine max
US-09-903-814-20

Query Match 59.3%; Score 1030; DB 9; Length 242;
Best Local Similarity 80.5%; Pred. No. 2.8e-98;
Matches 194; Conservative 22; Mismatches 25; Indels 0; Gaps 0;

QY 52 KKKACEAVGKSYEVNLPDESSEDEVLKHIAFNSDPSVHGILVOLPLPHMNDENILNA 111
Db 2 EKKACESVGINSLEANLPENSTEEVLYAGYNDPSVHGILVOLPLPHMNEQNILNA 61
QY 112 VSTEKDVGDGPHPLNIGRLAMQGRDFFVPCTPKGCMELHRYGVEIKGRAVIVGRSNIIV 171

Db 62 VRIEKVDGPHPLNIGRLAMRGREPLFVPCPTKGCIBLLHRYNVSIGKRAVVGSRNIV 121
QY 172 GMPAALLQKANATVSVHSNTKPEETITROADIVIAAVGVANLVRSWKPGAAITDVG 231
Db 122 GMPAALLQREDAIVSVHSNTSPETIIRQADIIIAAVQANVRSWKPGAVIIDVG 181
QY 232 INPVDDPESPRGYRLGVDVCYEEASKIAGLITPVGGVGPMTIAMLSNTLESARKHKF 291
Db 182 INPVDDPESPRGYRLGVDVCYEEAIRASAVTPVGGVGPMTIAMLQNTLISAKRVHNF 241
QY 292 K 292
Db 242 E 242

RESULT 5

US-10-723-061-20
; Sequence 20, Application US/10723061
; Publication No. US20040132150A1

; GENERAL INFORMATION:

; APPLICANT: Falco, Carl
; APPLICANT: Farnodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV1
; CURRENT APPLICATION NUMBER: US/10723,061
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: US 09/903,814
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Glycine max

US-10-723-061-20

Query Match 69.3%; Score 1030; DB 16; Length 242;
Best Local Similarity 80.5%; Pred. No. 2.8e-98;
Matches 194; Conservative 22; Mismatches 25; Indels 0; Gaps 0;

QY 52 KKKACEAVGHSYEVNLPEDSSEDEVLKHATFNSDPSVHGILVQLPLPHHMDENILNA 111
Db 2 EKKACESVGINSLEANLPENSTEEVLNYIAGYNDPSVHGILVQLPLPSHMNEQILNA 61
QY 112 VSIKDVGDGPHPLNIGRLAMRGREPLFVPCPTKGCIBLLHRYNVSIGKRAVVGSRNIV 171
Db 62 VRIEKVDGPHPLNIGRLAMRGREPLFVPCPTKGCIBLLHRYNVSIGKRAVVGSRNIV 121
QY 172 GMPAALLQKANATVSVHSNTKPEETITROADIVIAAVGVANLVRSWKPGAAITDVG 231
Db 122 GMPAALLQREDAIVSVHSNTSPETIIRQADIIIAAVQANVRSWKPGAVIIDVG 181
QY 232 INPVDDPESPRGYRLGVDVCYEEASKIAGLITPVGGVGPMTIAMLSNTLESARKHKF 291
Db 182 INPVDDPESPRGYRLGVDVCYEEAIRASAVTPVGGVGPMTIAMLQNTLISAKRVHNF 241
QY 292 K 292
Db 242 E 242

RESULT 6

US-10-425-114-60605
; Sequence 60605, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 60605
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3597-067-D5_FLI.pep
US-10-425-114-60605

Query Match 69.0%; Score 1025; DB 12; Length 292;
Best Local Similarity 66.9%; Pred. No. 1.2e-97;
Matches 192; Conservative 42; Mismatches 53; Indels 0; Gaps 0;

QY 1 AKTIIDCKLVAKQIRBEIAVEIAKMKDATGVVGLAVILVGSRKDSQTYVYRNKKKCAEAVG 60
Db 2 AQIIDGKAIAADVREVAADVAALSAHGLVPLGAVIVGSRKDSQTYVYRNKKKCAEAVG 61
QY 61 IKSXEYVNLPEDSSEDEVLKHATFNSDPSVHGILVQLPLPHHMDENILNAVSIKDVGD 120
Db 62 ICSIDVDLPEDISETALVAEVRHLNADPAVHGILVQLPLPKHINEKILSEISIEKDVGD 121
QY 121 FHPLNIGRLAMOGRDPPFVPCPTKGCIMELLHRYGVEIKGKRAVVGSRNIVGMPAALLQ 180
Db 122 FHPLNIGRLAMGRBFLFVPCPTKGCIMELLSRGVTVKGRVAVVGRSNIVGLPSVLL 181
QY 181 KANATVSVHSNTKPEETITROADIVIAAVGVANLVRSWKPGAAIIDVGINPVDPPS 240
Db 182 KADATVSVHSNTPDPESIVREADIVIAAGQAMWIKGDIKFGAVIDVGTNSVDDPTR 241
QY 241 PRGYRLGVDVCYEEASKIAGLITPVGGVGPMTIAMLSNTLESARK 287
Db 242 KSGYRLGVDVDFAAASKVAGYLTPTVGGVGPMTVAMLLKNTVDDGAKR 288

RESULT 7

US-10-425-114-56838
; Sequence 56838, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 56838
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLM017055A08_FLI.pep
US-10-425-114-56838

Query Match 69.0%; Score 1025; DB 12; Length 299;
Best Local Similarity 66.9%; Pred. No. 1.3e-97;
Matches 192; Conservative 42; Mismatches 53; Indels 0; Gaps 0;


```

: GENERAL INFORMATION:
: APPLICANT: La Rosa,Thomas J
: APPLICANT: Kovalic,David K
: APPLICANT: Zhou Yihua
: APPLICANT: Cao Yongwei
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 39-21(53223)B
: CURRENT APPLICATION NUMBER: US/10/424,599
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 268535
: LENGTH: 294
: TYPE: PRT

```

ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_8450C.1.pap
US-10-424-599-268535

Query Match 68.8%; Score 1010; DB 12; Length 294;
Best Local Similarity 65.5%; Pred. No. 4.5e-96;
Matches 186; Conservative 47; Mismatches 52; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 60
DB 2 ATVIDGKVAQIIRSEIADEVRLLSQKYGVKVGVLAVVIGNRKDSQSYVGMKRCACAEIG 61
QY 61 IKSYEVNLPDSSSEDEVLKHIAATFNSDPSVHGILVOLPLPHMNDENILNAVSIKXVDG 120
DB 62 IKSFVDLPEQVSQAEILIKQVHNLNPNPDVHGILVOLPLPKHINEBEVLTEISLEKXVDG 121
QY 121 FHLNIGRLAMQGRDPFFVPCPKGMELHRYGVVEIKGRAVIGRSNIVGMPAALLQ 180
DB 122 FHLNIGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 181
QY 181 KANATSVIHSNTKKPEETRQADIVIAAGVANLVGRSWIKPGAAIIDVGINPDDPS 240
DB 182 KADATVTIHSHTSQPESIIHEADIVIAAGQPMIKGNWIKPMIKNKIKGAAVIDVGTNAVDDPTK 241
QY 241 PRGYRLVGDVGYBEASKIAGLITPVGCGVCPMTIAMLLSNTLESAR 287
DB 242 KSGYRLVGDVDFGEASKVAGWITPVGCGVCPMTIAMLLSNTLESAR 288

RESULT 14

US-10-425-114-44200
Sequence 44200, Application US/10425114
Publication No. US20040034898A1

GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 44200
LENGTH: 321
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 700847483_FLI.pap
US-10-425-114-44200

Query Match 66.9%; Score 994; DB 12; Length 321;
Best Local Similarity 65.2%; Pred. No. 2.4e-94;
Matches 187; Conservative 45; Mismatches 55; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 60
DB 29 ATVIDGKVAQIIRSEIADEVRLLSQKYGVKVGVLAVVIGNRKDSQSYVGMKRCACAEIG 88
QY 61 IKSYEVNLPDSSSEDEVLKHIAATFNSDPSVHGILVOLPLPHMNDENILNAVSIKXVDG 120
DB 89 IKSFVDLPEQVSQAEILIKQVHNLNPNPDVHGILVOLPLPKHINEBEVLTEISLEKXVDG 148
QY 121 FHLNIGRLAMQGRDPFFVPCPKGMELHRYGVVEIKGRAVIGRSNIVGMPAALLQ 180
DB 149 FHLNIGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 208
QY 181 KANATSVIHSNTKKPEETRQADIVIAAGVANLVGRSWIKPGAAIIDVGINPDDPS 240

DB 209 KADATVTIHSHTSQPESIIHEADIVIAAGQPMIKGNWIKPMIKNKIKGAAVIDVGTNAVDDPTK 268
QY 241 PRGYRLVGDVGYBEASKIAGLITPVGCGVCPMTIAMLLSNTLESAR 287
DB 269 KSGYRLVGDVDFGEASKVAGWITPVGCGVCPMTIAMLLSNTLESAR 315

RESULT 15

US-10-437-963-125501
Sequence 125501, Application US/10437963
Publication No. US20040123343A1

GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 125501
LENGTH: 395
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_28138C.1.pap
US-10-437-963-125501

Query Match 66.5%; Score 988; DB 16; Length 395;
Best Local Similarity 65.0%; Pred. No. 1.4e-93;
Matches 186; Conservative 45; Mismatches 55; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 60
DB 98 ATVIDGKVAQIIRSEIADEVRLLSQKYGVKVGVLAVVIGNRKDSQSYVGMKRCACAEIG 157
QY 61 IKSYEVNLPDSSSEDEVLKHIAATFNSDPSVHGILVOLPLPHMNDENILNAVSIKXVDG 120
DB 158 IKSLAELPGNCTEDVVDVSVSRFNEPDSVHGILVOLPLPKHINEBEVLTEISLEKXVDG 217
QY 121 FHLNIGRLAMQGRDPFFVPCPKGMELHRYGVVEIKGRAVIGRSNIVGMPAALLQ 180
DB 218 FHLNIGKLVAKQIREIAVEIAAKMDAIGVPGVLAVILVGRKDSQTYVYRNKKKACEAVG 277
QY 181 KANATSVIHSNTKKPEETRQADIVIAAGVANLVGRSWIKPGAAIIDVGINPDDPS 240
DB 278 RHATVSIHAFITNPPEETRQSDIVISAAGVANLVGRSWIKPGAAIIDVGINPDDPS 337
QY 241 PRGYRLVGDVGYBEASKIAGLITPVGCGVCPMTIAMLLSNTLESAR 286
DB 338 DYGRLVGDVDFGEASKVAGWITPVGCGVCPMTIAMLLSNTLESAR 383

Search completed: July 24, 2004, 07:09:11
Job time : 970 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:34:25 ; Search time 57 Seconds
(without alignments)
264.470 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIDGKLVAKQIREEIAVE.....TIAMLLSNTLESAKRIHKFK 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pdp.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pdp.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pdp.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pdp.*
- 5: /cgm2_6/ptodata/2/iaa/PTCTUS_COMB.pdp.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1486	100.0	292	4	US-09-903-814A-18
2	1030	69.3	242	4	US-09-903-814A-20
3	1016	68.4	290	4	US-09-903-814A-16
4	712.5	47.9	284	4	US-09-107-532A-7288
5	701	47.2	289	4	US-09-903-814A-22
6	684	46.0	318	4	US-09-489-039A-13257
7	679.5	45.7	290	4	US-09-134-001C-3401
8	677	45.6	318	4	US-09-252-991A-30283
9	663.5	44.7	280	4	US-09-134-000C-3865
10	661	44.5	290	4	US-09-543-681A-7078
11	623.5	42.0	344	3	US-09-318-448-21
12	608.5	40.9	314	4	US-09-328-352-5492
13	518	34.9	271	4	US-09-198-452A-353
14	104.5	7.0	482	4	US-09-252-991A-19130
15	99.5	6.7	231	4	US-08-426-630-54
16	98	6.6	382	4	US-09-134-001C-3765
17	92	6.2	1025	4	US-09-711-164-443
18	90.5	6.1	409	4	US-09-134-001C-3790
19	88	5.9	202	4	US-08-311-731A-13
20	85.5	5.8	628	4	US-09-252-991A-32414
21	85.5	5.8	650	4	US-09-252-991A-27296
22	83.5	5.6	648	3	US-09-183-706-43
23	83.5	5.6	648	4	US-09-567-995-34
24	83.5	5.6	865	4	US-09-711-164-354
25	83	5.6	300	4	US-09-328-352-6714
26	82.5	5.6	432	4	US-09-489-039A-13935
27	82.5	5.6	524	4	US-09-540-236-2706

28	82	5.5	502	4	US-09-134-001C-3598	Sequence 3598, Ap
29	82	5.5	549	3	US-08-886-886-13	Sequence 13, Appl
30	82	5.5	549	4	US-09-134-000C-5876	Sequence 5876, Ap
31	81.5	5.5	262	4	US-09-134-000C-4224	Sequence 4224, Ap
32	81.5	5.5	430	4	US-08-311-731A-153	Sequence 153, Appl
33	81.5	5.5	664	4	US-09-252-991A-30117	Sequence 30117, A
34	81	5.5	347	4	US-09-107-532A-6054	Sequence 6054, Ap
35	81	5.5	520	1	US-08-706-292-2	Sequence 2, Appl
36	81	5.5	520	3	US-09-032-365A-15	Sequence 15, Appl
37	81	5.5	904	4	US-09-543-681A-4485	Sequence 4485, Ap
38	80.5	5.4	357	3	US-09-116-115-17	Sequence 17, Appl
39	80.5	5.4	357	4	US-09-541-762-17	Sequence 17, Appl
40	80	5.4	680	4	US-09-717-364A-19	Sequence 19, Appl
41	80	5.4	924	4	US-09-267-311-2	Sequence 2, Appl
42	79.5	5.3	269	4	US-09-252-991A-27159	Sequence 27159, A
43	79.5	5.3	523	2	US-08-482-728A-19	Sequence 19, Appl
44	79.5	5.3	523	3	US-09-028-366-4	Sequence 4, Appl
45	79.5	5.3	523	4	US-09-715-285-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-903-814A-18
; Sequence 18, Application US/099030814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Ramodu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 292
; TYPE: PRI
; ORGANISM: Oryza sativa
US-09-903-814A-18

Query Match 100.0%; Score 1486; DB 4; Length 292;
Best Local Similarity 100.0%; Pred. No. 8.5e-165;
Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	AKIDGKLVAKQIREEIAVEIAKMDAIGVVGGLAVILVGSRKDSQTYVRNKKCAEAVG	60
Db	1	AKIDGKLVAKQIREEIAVEIAKMDAIGVVGGLAVILVGSRKDSQTYVRNKKCAEAVG	60
Qy	61	IKSEYVNLDESDSEDLVKHATFNSDPSVHGILVQLPLPHMNDENILNAVSIKDVGD	120
Db	61	IKSEYVNLDESDSEDLVKHATFNSDPSVHGILVQLPLPHMNDENILNAVSIKDVGD	120
Qy	121	FHPLNIGRLAMQGRDPFFVPCPKGMELHRYGVEIKGKRAVWIGRSNIVGMPAALLIQ	180
Db	121	FHPLNIGRLAMQGRDPFFVPCPKGMELHRYGVEIKGKRAVWIGRSNIVGMPAALLIQ	180
Qy	181	KANATYSIVHSNTKPEEITROADIVIAAGVANLVGSGWIKPGAAIDVGINPVDPPS	240
Db	181	KANATYSIVHSNTKPEEITROADIVIAAGVANLVGSGWIKPGAAIDVGINPVDPPS	240
Qy	241	PRGYRLVGVGVCEEASAKIAGLITPVGPGVGPMTIAMLLSNTLESAKRIHKFK	292
Db	241	PRGYRLVGVGVCEEASAKIAGLITPVGPGVGPMTIAMLLSNTLESAKRIHKFK	292

```
RESULT 2
US-09-903-814A-20
; Sequence 20, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 20
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Glycine max
; US-09-903-814A-20

Query Match          69.3%; Score 1030; DB 4; Length 242;
Best Local Similarity 80.5%; Pred. No. 1.1e-111;
Matches 194; Conservative 22; Mismatches 25; Indels 0; Gaps 0;

QY 52 KKACBAVGKSVENLPEDSSDEVLKHIAITFNSDPVHGILVOLPLPHHNDENILNA 111
DB 2 EKKACBSVGSINLEANLPENSTEEVLYNTAGYNDPSVHGILVOLPLPSHNEQILNA 61

QY 112 VTEKVDGPHPLNIGRLAMQGRDPFVPCPKGCMELLHRYGVETKGRVAVIGRSNV 171
DB 62 VRIEKVDGPHPLNIGRLAMQGRDPFVPCPKGCMELLHRYGVETKGRVAVIGRSNV 121

QY 172 GMPAALLQKANATVSIHNSNTKPEIITRQADIVIAAVGVANLVRSWIKPGAAIIVDG 231
DB 122 GMPAALLQKEDATVSIHNSNTKPEIITRQADIVIAAVGVANLVRSWIKPGAAIIVDG 181

QY 232 INPVDPESPRGVRLVGVYEEASIKAGLITPVGCGVGPMTIAMLLSNTLESARKHF 291
DB 182 INPVDPESPRGVRLVGVYEEASIKAGLITPVGCGVGPMTIAMLLSNTLESARKHF 241

QY 292 K 292
DB 242 E 242

RESULT 3
US-09-903-814A-16
; Sequence 16, Application US/09903814A
; Patent No. 6680428
; GENERAL INFORMATION:
; APPLICANT: Falco, Carl
; APPLICANT: Famodu, Lavo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; APPLICANT: Thorpe, Cathy
; TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
; FILE REFERENCE: BB1179 USDIV
; CURRENT APPLICATION NUMBER: US/09/903,814A
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: US 60/092,869
; PRIOR FILING DATE: 1998-07-15
; PRIOR APPLICATION NUMBER: US 09/351,703
; PRIOR FILING DATE: 1999-07-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16

; LENGTH: 290
; TYPE: PRT
; ORGANISM: Zea mays
US-09-903-814A-16

Query Match          68.4%; Score 1016; DB 4; Length 290;
Best Local Similarity 66.2%; Pred. No. 6.3e-110;
Matches 190; Conservative 43; Mismatches 54; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAQIREETAVETAKDAIGVVGGLAVILGSRKDSQTYVYNNKKACEAVG 60
DB 2 AQIIDGKAIADVRREVAADVAALSAHGLVPLGVVIVGSRKDSQTYVYNNKKACEAVG 61

QY 61 KSYEVNLPEDSSDEVLKHIAITFNSDPVHGILVOLPLPHHNDENILNAVSIKEDVDG 120
DB 62 ICSIDVDLPEDISETALVAHVRLNADPAVHGLVOLPLPKHINEKILSEISIEKVDVG 121

QY 121 FHLNIGRLAMQGRDPFVPCPKGCMELLHRYGVETKGRVAVIGRSNVGMPAALLQ 180
DB 122 FHLNIGRLAMQGRDPFVPCPKGCMELLHRYGVETKGRVAVIGRSNVGMPAALLQ 181

QY 181 KANATVSIHNSNTKPEIITRQADIVIAAVGVANLVRSWIKPGAAIIVDGINPVDPE 240
DB 182 KANATVSIHNSNTKPEIITRQADIVIAAVGVANLVRSWIKPGAAIIVDGINPVDPE 241

QY 241 PRGYRLVGVYEEASIKAGLITPVGCGVGPMTIAMLLSNTLESARK 287
DB 242 KSGVRLVGVYEEASIKAGLITPVGCGVGPMTIAMLLSNTLESARK 288

RESULT 4
US-09-107-532A-7288
; Sequence 7288, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Dereke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: STC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 7288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
```

```
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...284
SEQUENCE DESCRIPTION: SEQ ID NO: 7288:
US-09-107-532A-7288

Query Match      47.9%; Score 712.5; DB 4; Length 284;
Best Local Similarity 51.0%; Pred. No. 1.7e-74;
Matches 146; Conservative 54; Mismatches 77; Indels 9; Gaps 5;

Qy 1 AKIIDGKLVAKQIREIEIAVEIAEMKDAIGVVPGLAVILVGRSKDSQTYVNRKKACEAVG 60
Db 2 AELINGKELAEKQAEIAEIKQLKDN-GIHPGLVLLVGENPASQIYVNRKVAKEIG 60
Qy 61 IKSYEVNLPEDSSDEVLKHIATFNSDPSVGHILVQLPLPHMNDENILNAVSIKDVGD 120
Db 61 IHSLVRYPTETISEDELLAEIEKNQDPSFPHGILVQLPLPKHIDEEKILLAIRPEKDVGD 120
Qy 121 FPHLNIHGLAMQGRDPFPVCTPKGCMELLHRYGVEIKGRAVIGRSNIVGMPAALLQ 180
Db 121 FPHNMLGL-LAGK-PDKIPTPGIMKMFAYSIDPAGRAVIGRSNIVGKPMQALL 178
Qy 181 KANATVSIHNSNTKPEBITRQADIVIAAVGVANLVGRSWIKPGAAIIDVGINPVDPPES 240
Db 179 MADATVTVHASKTENLAELAEADILVVAIGRGHFVTKFVKPGAVVIDVGMN----RD 233
Qy 241 PRGYRLVGDVCEYEASKIAGLITPVGGVGWPTTAMLSNTLESAS 286
Db 234 ANG-KLIGDVRFDEVEPIAGWITPVKGVGPTTITMLMYQIVKSAE 278

RESULT 5
US-09-903-814A-22
Sequence 22, Application US/09903814A
Patent No. 6680428
GENERAL INFORMATION:
APPLICANT: Falco, Carl
APPLICANT: Farnold, Lavo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
APPLICANT: Thorpe, Cathy
TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
FILE REFERENCE: BB1179 USDI
CURRENT APPLICATION NUMBER: US/09/903.814A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: 2001-07-12
PRIOR FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: US 60/092,869
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: US 09/351,703
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Microsoft Office 97
SEQ ID NO 22
LENGTH: 289
TYPE: PRT
ORGANISM: Triticum aestivum
FEATURE:
NAME/KEY: UNSURE
LOCATION: (195)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (227)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (233)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (235)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
```

```
FEATURE:
NAME/KEY: UNSURE
LOCATION: (252)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (256)..(257)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (260)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (267)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (272)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (278)..(279)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (281)..(282)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
FEATURE:
NAME/KEY: UNSURE
LOCATION: (286)
OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-903-814A-22
```

```
Query Match      47.2%; Score 701; DB 4; Length 289;
Best Local Similarity 58.4%; Pred. No. 3.8e-73;
Matches 135; Conservative 37; Mismatches 59; Indels 0; Gaps 0;

Qy 1 AKIIDGKLVAKQIREIEIAVEIAEMKDAIGVVPGLAVILVGRSKDSQTYVNRKKACEAVG 60
Db 7 AQIIDGKATAAEIRREIGAEVAVLS9AHNIVFGLAVIVGSRKDSQTYVQMKRKAQAEVG 66
Qy 61 IKSYEVNLPEDSSDEVLKHIATFNSDPSVGHILVQLPLPHMNDENILNAVSIKDVGD 120
Db 67 IRSFDVLDPEIDAEALVAEVELNADPAVHGLVQLPLPKHINEENILNQISIEKDVGD 126
Qy 121 FPHLNIHGLAMQGRDPFPVCTPKGCMELLHRYGVEIKGRAVIGRSNIVGMPAALLQ 180
Db 127 FPHLNIHGLAMQGRDPFPVCTPKGCMELLHRYGVEIKGRAVIGRSNIVGMPAALLQ 186
Qy 181 KANATVSIHNSNTKPEBITRQADIVIAAVGVANLVGRSWIKPEGAALIDVG 231
Db 187 ESGRYVDXASTDPNPTISVKQDIVIAAGQAMMIKGDWKKQKRNKXIXVG 237
```

```
RESULT 6
US-09-489-039A-13257
Sequence 13257, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Bregon et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 13257
LENGTH: 318
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
```



```
QY 3 IIDGKLVAKQIREIEIAEIAKMDAIGVPGGLAVILVGSRKDSQTYVRNKKKACEAVGK 62
Db 12 VINGRELADQMAEIQKDVKNMTQ-QIQPLGVLVLLVGENPASQTYVRNKKKAAKIGIL 70
QY 63 SYEVNLPEDSSEDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDVDPGH 122
Db 71 SKVEKLPTTISEEELALBIDINQDSRPHGLVQPLPKHIDEEKILLALIDPKDVGDFH 130
QY 123 PLNIGRLAMQGRDPFFVPCPKGCMELLHRYGVEIKGKRAVIVGRSNIVGMPAALLQKA 182
Db 131 PMNLGRLV-GK-PEMIPCTPVGIMKMEAYDIDLTKRAVIVGRSNIVGKEMAQLLLMK 188
QY 183 NATVISVHSNTKKPEEITRQADIVIAAVGVANLVGSGWIKKPGAAIIDVGINPVDPSR 242
Db 189 NATVIAHSKTEHLAEVAKEADILVAIGRHFVTKFEVKGAVVIDVGMN-----RQCE 243
QY 243 GYRLVGDVGYEASKIAGLITPVPGGVGPM 272
Db 244 G-KLIGDAFVDEVSIAISYIIPVPGGVGPM 272

RESULT 10
US-09-543-681A-7078
; Sequence 7078, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7078
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7078

Query Match 44.5%; Score 661; DB 4; Length 290;
Best Local Similarity 48.3%; Pred. No. 1.8e-68;
Matches 140; Conservative 47; Mismatches 93; Indels 10; Gaps 4;

QY 1 AKIIDGKLVAKQIREIEIAEIAKMDAIGV-VPGGLAVILVGSRKDSQTYVRNKKKACEAV 59
Db 3 ARIIDGKTARTIRSEVA-EKVQRIKIGKRAPGLAVILVGNPASQIYVASKRACDEV 61
QY 60 GIKSYEVNLPEDSSEDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDV 119
Db 62 GFISRYDLPDTTSEADLNLDTLNEEDNTIDGILVQLPLPAGIDNVKVLRIHPDKVD 121
QY 120 GHPLNIGRLAMQGRDPFFVPCPKGCMELLHRYGVEIKGKRAVIVGRSNIVGMPAALL 179
Db 122 GHYPNIGRLCQ--RAPKLRCPTPRGIVTLERCNIPMNLNGLNAVIGASNIVGRPMSLE 179
QY 180 QKANATVISVHSNTKKPEEITRQADIVIAAVGVANLVGSGWIKKPGAAIIDVGINPVD 239
Db 180 LLAGCTTTVTHRTKDLRHFVEHADLVVAVGKFNIPGEWIKPGAAIVDVGINLRENG 238
QY 240 SPFGYRLVGDVGYEASKIAGLITPVPGGVGPMTHAMLLSNTLESKRIH 289
Db 239 -----KVVGVDVPAQASQSRAGNISPVPGGVGPMVTATLQNTLOACEYH 283

RESULT 11
US-09-318-448-21
; Sequence 21, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
```

```
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 21
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-318-448-21

Query Match 42.0%; Score 623.5; DB 3; Length 344;
Best Local Similarity 46.2%; Pred. No. 5.6e-64;
Matches 138; Conservative 52; Mismatches 94; Indels 15; Gaps 5;

QY 1 AKIIDGKLVAKQIREIEIAEIAKMDAIGVPGGLAVILVGSRKDSQTYVRNKKKACEAVG 60
Db 31 AVVISGRKLAQIKQEVROEVEEWSASGNKRPHLSVILVGENPASHSYVLNKTAAAAVVG 90
QY 61 IKSYEVNLPEDSSEDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDV 120
Db 91 INSETIMKPAISSEBELNLINKLNNDNDVGLVQLPLPEHIDERRICNAVSPDKDVG 150
QY 121 FHLNIGRLAMQGRDPF-FVPCPKGCMELLHRYGVEIKGKRAVIVGRSNIVGMPAALL 179
Db 151 FHVINGRMLC---DOYSMLPATPWGWEIITKGTIGTLGKVVVAGRSKNVGMPIAMLL 207
QY 180 QK-----ANATVISVHSNTKKPE--EITROADIVIAAVGVANLVGSGWIKKPGAA 229
Db 208 HTDGAHERPGGDATVITSHRYTPKEQLKHTILADIVISAAGIPNLITADMIKEGAA 267
QY 230 VGINPVDPSPRGYRLVGDVGYEASKIAGLITPVPGGVGPMTHAMLLSNTLESKRI 288
Db 268 VGINRVHDPVTAKP-KLVGVDVDFEGVQKAGYITPVPGGVGPMTHAMLMKNTIAAK 325

RESULT 12
US-09-328-352-5492
; Sequence 5492, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5492
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5492

Query Match 40.9%; Score 608.5; DB 4; Length 314;
Best Local Similarity 43.2%; Pred. No. 2.7e-62;
Matches 124; Conservative 56; Mismatches 98; Indels 9; Gaps 2;

QY 1 AKIIDGKLVAKQIREIEIAEIAKMDAIGVPGGLAVILVGSRKDSQTYVRNKKKACEAV 60
Db 34 ALVLDGRLAKQIREIEIAEIAKMDAIGVPGGLAVILVGSRKDSQTYVRNKKKACEAV 93
QY 61 IKSYEVNLPEDSSEDEVILKHIATFNSDPSVHGILVOLPLPHMNDENILNAVSIKDV 120
Db 94 MDSLKIELPQETTEQLLAETIEKLNANPVDVHGILQHPVPAQIDRACFDALSLAKDVG 153
QY 121 FHLNIGRLAMQGRDPF-FVPCPKGCMELLHRYGVEIKGKRAVIVGRSNIVGMPAALL 180
Db 154 VTCJGFGRMV--GEAAYGATPAGIMTILKENNIEIAGKHAVVVGSAIGKPMANMLL 211
QY 181 KANATVISVHSNTKKPEEITRQADIVIAAVGVANLVGSGWIKKPGAAIIDVGINPVD 240
```

Query Match 34.9%; Score 518; DB 4; Length 271;
Best Local Similarity 41.5%; Pred. No. 7.9e-52;
Matches 113; Conservative 54; Mismatches 87; Indels 18; Gaps 6;
Db 212 QANATVTCISRTQNLVELVQADIIIVGAVGKAEIQRDKIKQAVVVDAGFHPDGGG- 270
Qy 241 PRGVLGVGVYBASKIAGITVPGGVGPMITAMLLSNTLESKR 287
Db 271 -----VGDIOLOIEEIASATVPGGVGPMITITLIRQTVAAEK 311
RESULT 13
US-09-198-452A-353
; Sequence 353, Application US/09.98452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
; TITLE OF INVENTION: 9710-003-999
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198.452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 353
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-353

Query Match 34.9%; Score 518; DB 4; Length 271;
Best Local Similarity 41.5%; Pred. No. 7.9e-52;
Matches 113; Conservative 54; Mismatches 87; Indels 18; Gaps 6;
Db 212 QANATVTCISRTQNLVELVQADIIIVGAVGKAEIQRDKIKQAVVVDAGFHPDGGG- 270
Qy 241 PRGVLGVGVYBASKIAGITVPGGVGPMITAMLLSNTLESKR 287
Db 271 -----VGDIOLOIEEIASATVPGGVGPMITITLIRQTVAAEK 311
RESULT 13
US-09-198-452A-353
; Sequence 353, Application US/09.98452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
; TITLE OF INVENTION: 9710-003-999
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198.452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 353
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-353

Query Match 34.9%; Score 518; DB 4; Length 271;
Best Local Similarity 41.5%; Pred. No. 7.9e-52;
Matches 113; Conservative 54; Mismatches 87; Indels 18; Gaps 6;
Db 212 QANATVTCISRTQNLVELVQADIIIVGAVGKAEIQRDKIKQAVVVDAGFHPDGGG- 270
Qy 241 PRGVLGVGVYBASKIAGITVPGGVGPMITAMLLSNTLESKR 287
Db 271 -----VGDIOLOIEEIASATVPGGVGPMITITLIRQTVAAEK 311
RESULT 13
US-09-198-452A-353
; Sequence 353, Application US/09.98452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
; TITLE OF INVENTION: 9710-003-999
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198.452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 353
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-353

Query Match 7.0%; Score 104.5; DB 4; Length 482;
Best Local Similarity 22.1%; Pred. No. 0.0039;
Matches 54; Conservative 39; Mismatches 72; Indels 79; Gaps 10;
Qy 29 GVVPLAVILVGSRKDSQTYVYRNKKKACEAVGKSYFVNLDPDSS-----EDEVL-- 78
Db 25 GLAGGVAQVL-----HAQAQVASRSHQFDAVALEAVLHLHLSGSAVAFAGFADELAL 79
Qy 79 -KHIAATFNSDPSVHGILVQLPLPHHMDENTINAVSIEKVDGPHPLNIGRLAMGQDPF 137
Db 80 GQMA-----VLVVLVWVHVAQAALGVAAGVGLGVHQAVERLA--GR--- 124
Qy 138 FVPCPTKGMCELLHRYGVEIKRAVVTGRNIVGMPAALLLQKANATVSVHNTKKPE 197
Db 125 -----RVVDRLAVDLGGARHW-----VGFGAALDLQR-----VHADPQPL 161
Qy 198 EITROADIV-IAAVGVANLVRGSMIKPGAAIIDVGINPDDPESPRGYRLVGDVCEBAS 256
Db 162 DVLHGAELVGLHDVGAVFLEG-----GHRFAGTVAFLEEE 197
Qy 257 KIAG 260
Db 198 HLLG 201

RESULT 15
US-08-426-630-54
; Sequence 54, Application US/08426630
; Patent No. 6656709
; GENERAL INFORMATION:
; APPLICANT: BLANCHE, FRANCIS; CAMERON, BEATRICE; CROUZET,
; APPLICANT: JOEL; DEBUSCHE, LAURENT; LEVY SCHIL, SOPHIE;
; APPLICANT: THIBAUT, DENIS
; TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE
; TITLE OF INVENTION: BIOSYNTHESIS OF COBALAMINS AND/OR COBAMIDES, DNA SEQUENCES
; TITLE OF INVENTION: CODING FOR THESE POLYPEPTIDES, PREPARATION METHOD AND THEIR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 555 13TH STREET, N.W.
; CITY: WASHINGTON
; STATE: DISTRICT OF COLUMBIA
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,630
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,151
; FILING DATE: 14-SEP-1992
; APPLICATION NUMBER: PCT/FR91/00054
; FILING DATE: 30-JAN-1991
; ATTORNEY FIRM INFORMATION:
; NAME: F. F. CALVETII
; REGISTRATION NUMBER: 28,557
; REFERENCE/DOCKET NUMBER: 1290-7213
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 857-7887
; TELEFAX: (202) 857-7929
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Linear

RESULT 15
US-08-426-630-54
; Sequence 54, Application US/08426630
; Patent No. 6656709
; GENERAL INFORMATION:
; APPLICANT: BLANCHE, FRANCIS; CAMERON, BEATRICE; CROUZET,
; APPLICANT: JOEL; DEBUSCHE, LAURENT; LEVY SCHIL, SOPHIE;
; APPLICANT: THIBAUT, DENIS
; TITLE OF INVENTION: POLYPEPTIDES INVOLVED IN THE
; TITLE OF INVENTION: BIOSYNTHESIS OF COBALAMINS AND/OR COBAMIDES, DNA SEQUENCES
; TITLE OF INVENTION: CODING FOR THESE POLYPEPTIDES, PREPARATION METHOD AND THEIR
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 555 13TH STREET, N.W.
; CITY: WASHINGTON
; STATE: DISTRICT OF COLUMBIA
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,630
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,151
; FILING DATE: 14-SEP-1992
; APPLICATION NUMBER: PCT/FR91/00054
; FILING DATE: 30-JAN-1991
; ATTORNEY FIRM INFORMATION:
; NAME: F. F. CALVETII
; REGISTRATION NUMBER: 28,557
; REFERENCE/DOCKET NUMBER: 1290-7213
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 857-7887
; TELEFAX: (202) 857-7929
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: Amino acid
; STRANDEDNESS:
; TOPOLOGY: Linear

MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Methanobacterium ivanovi
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
FEATURE:
NAME/KEY: CORA
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: Translation product of SEQ ID NO:53
US-08-426-630-54

Query Match 6.7%; Score 99.5; DB 4; Length 231;
Best Local Similarity 21.3%; Pred No. 0.0044;
Matches 56; Conservative 46; Mismatches 94; Indels 67; Gaps 12;

QY	34	LAVILVGRKDSQTYVRNKKACEAVGI-KSYEYNLPEDSSEDEVLKHIAFNSDPSVHG	92
Db	1	MVYLVGAGPGDPPELITLK-----AVNLKKADVVLVDKPAEEILKYA-----EGAKL	49
QY	93	ILVQLPLPHMNDENILNAVSTIEKDVDFGHPINIGRLAMQGRDPFFVPCPKGCMELH-	151
Db	50	IYVGKQAGHYKXQNEINTLLVEEAXE-----NDLVRLKGDPEFV---GRGGEIILAL	101
QY	152	RYGVEIKGRVAVIGRNVIGMPAAL---LIQKANATVSIVHSNTKKPEITRQ-----	202
Db	102	VEEGIDFE-----LVPGVISAIGVPTTIGLPVTHRGVATSFVTWTHGEDPTCKKQVGWDF	157
QY	203	ADIVIAAVGVANLVGRGWIKPGAAIIDVGINPVDPPSPRGYRLVGDVCYEEASKIAGL	261
Db	158	KADTIVILMIGIGNLAENT-----AEIMKKKDPETP-----VCVIEN-----	193
QY	262	ITPVFGGVPMTIAMLLSNTLES	284
Db	194	-----GTMEGQRIITCTLEN	208

Search completed: July 24, 2004, 06:41:22
Job time : 59 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2004, 04:10:50 ; Search time 101 Seconds
(without alignments)
816.870 Million cell updates/sec

Title: US-10-723-061-18
Perfect score: 1486
Sequence: 1 AKIDGKLVAKQIREIAVE.....TIAMLLNTLESAKRIHKFK 292

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

ALIGNMENTS

RESULT 1
AAG29448
ID AAG29448 standard; protein; 357 AA.
XX AAG29448;
XX AC
XX 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 35041.
XX XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX OS Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX XX
XX 25-FEB-2000; 2000EP-00301439.
XX XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132863P.
XX 14-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134321P.
XX 14-MAY-1999; 99US-0134321P.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	83.3	357	3 AAG29448	Arabidopsis
2	1238	83.3	360	3 AAG29447	Arabidopsis
3	1238	83.3	426	3 AAG29446	Arabidopsis
4	1035	69.7	299	3 AAG22173	Arabidopsis
5	1035	69.7	299	3 AAG08973	Arabidopsis
6	1017	68.4	350	3 AAG36762	Arabidopsis
7	1017	68.4	352	3 AAG36761	Arabidopsis
8	1017	68.4	498	3 AAG36760	Arabidopsis
9	1012	68.1	350	3 AAG14693	Arabidopsis
10	1012	68.1	352	3 AAG14692	Arabidopsis
11	1012	68.1	377	3 AAG14691	Arabidopsis
12	762	51.3	934	4 AB60394	Drosophila
13	755.5	50.8	284	5 AB548308	Listeria
14	743.5	50.0	286	5 ABP25547	Streptococcus
15	739.5	49.8	946	6 ABP52676	Protein s
16	739.5	49.8	946	6 ABP57428	Saccharom
17	735.5	49.5	938	7 ABP76951	A. gossyp
18	735	49.5	816	4 AAG02709	Novel hum
19	724	48.7	285	6 ABU01214	S. pneumo
20	724	48.7	286	3 AAY81531	Streptococ
21	721	48.5	284	5 ABP25546	Streptococ
22	713.5	48.0	975	6 ABP52679	Protein s
23	712.5	47.9	284	7 ADC97661	E. faeciu
24	707.5	47.6	294	5 ABP54169	Lactococc
25	679.5	45.7	290	5 ABP38556	Staphyloc

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 03-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 08-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140855P.
PR 28-JUN-1999; 99US-0140855P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 05-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147038P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147302P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159395P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.

```
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161931P.
PR 28-OCT-1999; 99US-0161931P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      83.3%; Score 1238; DB 3; Length 357;
Best Local Similarity 79.1%; Pred. No. 1.9e-123;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

Qy 1 AKITDGLVAKQIREETAEVETAKMKDAIGVYVGLAVILVGRKDSQTYVRNKKACBAVG 60
Db 66 AIVIDGKAVAKIRDEITIEVSRNKEIGVIPGLAVILVGRKDSATYVRNKKACDSVG 125
Qy 61 KSYEVNLPEDSSDEVLKHIATFNPSVHGILVQLPLPHNMNDENILNAVIEKDVVG 120
Db 126 IKSEVRLAEDSSREEVLKSVGFNDPSVHGILVQLPLPSHMDQNILNAVSTERKDVVG 185
Qy 121 FHPNIGRLAQGRDPFVPCPTKGCVELLHRYGVEIKGRVAVIGRSNIVGMPAALLQ 180
Db 186 FHPNIGRLAMRGREPLFVPCPTKGCIELLHRYNIEIKGRVAVIGRSNIVGMPAALLQ 245
Qy 181 KANATSVIVHSNTKKPEITRQADIVIAAVGVANLVGRSMIKPGAAIDVGINPDDPES 240
Db 246 REDATVSIHSRTKNPEITREADIILISAVGQPNMVGSMIKPGAVLIDVGINPVEDPSA 305
Qy 241 PRGRLVGDVCIYEASKIAGLITVPGVGVPMTTAMLLSNLTLESARKIHFK 292
Db 306 ARGRLVGDICIEEASKVASAITVPGVGVPMTTAMLLSNLTLSAKRIHNFQ 357

RESULT 2
AAG29447
ID AAG29447 standard; protein; 360 AA.
AC AAG29447;
XX
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 35040.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0130891P.
XX 30-APR-1999; 99US-0131443P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
```

```
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134376P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141542P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
```

```
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 22-JUL-1999; 99US-0145218P.
PR 22-JUL-1999; 99US-0145224P.
PR 22-JUL-1999; 99US-0145276P.
PR 22-JUL-1999; 99US-0145313P.
PR 22-JUL-1999; 99US-0145318P.
PR 22-JUL-1999; 99US-0145319P.
PR 22-JUL-1999; 99US-0145351P.
PR 22-JUL-1999; 99US-0145386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 02-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-01493368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0158029P.
PR 07-OCT-1999; 99US-0158232P.
PR 08-OCT-1999; 99US-0158369P.
PR 12-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159320P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.

PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160800P.
PR 22-OCT-1999; 99US-0160801P.
PR 22-OCT-1999; 99US-0160899P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 83.3%; Score 1238; DB 3; Length 360;
Best Local Similarity 79.1%; Pred. No. 1.9e-123;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

QY 1 AKIIDGLVAKQIREIEIAVEIAKMDAIGVYVGLAVILVGRKDSQTYVRNKKKACEAYG 60
Db AIVIDGKAVAKIRDEITIEVSRMKESIGVPGILAVILVGRKDSATYVRNKKKACDSVG 128

QY 61 KSYEVNLPEDSSEDEVKHIATFNPSDPSVHGILVOLPLPHNMNDENILNAVSIKDVVG 120
Db IKSEVELAEDSEEEVSKSVGNFDDPSVHGILVOLPLPHNMNDENILNAVSIKDVVG 188

QY 121 FHPLNIGRLAMQGRDPFPVCTPKGCMELHRYGVEIKGKRAVIGRSNIVGMPAALLQ 180
Db FHPLNIGRLAMRGREPLFVCTPKGCIELHRYNIEIKGKRAVIGRSNIVGMPAALLQ 248

QY 181 KANATSVIHSNTKKPBEITRQADIVIAVGVANLVGRSVIKPFGAALIDVGINPVDPPS 240
Db REDATVSIHSRTKNPBEITREADIISAVGQNMVGRSVIKPFGAALIDVGINPVDPPSA 308

QY 241 PRGYRLVGDVCEASAKIAGLITPVGVPMTIAMLLSNTLESAKRIHKFK 292
Db ARGVRLVGDICYEASKVASAITPVGVPMTIAMLLSNTLTSAKRIHNFQ 360

RESULT 3
AAG29446
ID AAG29446 standard; protein; 426 AA.
XX AAG29446;
AC AAG29446;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 35039.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
```

PR	08-APR-1999;	99US-0128714P.	PR	19-JUL-1999;	99US-0144333P.
PR	16-APR-1999;	99US-0129845P.	PR	19-JUL-1999;	99US-0144334P.
PR	19-APR-1999;	99US-0130077P.	PR	19-JUL-1999;	99US-0144335P.
PR	21-APR-1999;	99US-0130443P.	PR	20-JUL-1999;	99US-0144352P.
PR	23-APR-1999;	99US-0130510P.	PR	20-JUL-1999;	99US-0144632P.
PR	23-APR-1999;	99US-0130891P.	PR	21-JUL-1999;	99US-0144884P.
PR	28-APR-1999;	99US-0131449P.	PR	21-JUL-1999;	99US-0144814P.
PR	30-APR-1999;	99US-0132048P.	PR	21-JUL-1999;	99US-0145088P.
PR	30-APR-1999;	99US-0132407P.	PR	21-JUL-1999;	99US-0145088P.
PR	04-MAY-1999;	99US-0132484P.	PR	22-JUL-1999;	99US-0145085P.
PR	05-MAY-1999;	99US-0132485P.	PR	22-JUL-1999;	99US-0145087P.
PR	06-MAY-1999;	99US-0132486P.	PR	22-JUL-1999;	99US-0145089P.
PR	06-MAY-1999;	99US-0132487P.	PR	22-JUL-1999;	99US-0145192P.
PR	07-MAY-1999;	99US-0132863P.	PR	23-JUL-1999;	99US-0145145P.
PR	11-MAY-1999;	99US-0134256P.	PR	23-JUL-1999;	99US-0145218P.
PR	14-MAY-1999;	99US-0134218P.	PR	23-JUL-1999;	99US-0145224P.
PR	14-MAY-1999;	99US-0134219P.	PR	26-JUL-1999;	99US-0145276P.
PR	14-MAY-1999;	99US-0134219P.	PR	27-JUL-1999;	99US-0145913P.
PR	14-MAY-1999;	99US-0134370P.	PR	27-JUL-1999;	99US-0145918P.
PR	18-MAY-1999;	99US-0134768P.	PR	27-JUL-1999;	99US-0145919P.
PR	19-MAY-1999;	99US-0134941P.	PR	28-JUL-1999;	99US-0145951P.
PR	20-MAY-1999;	99US-0135124P.	PR	02-AUG-1999;	99US-0146386P.
PR	21-MAY-1999;	99US-0135353P.	PR	02-AUG-1999;	99US-0146388P.
PR	21-MAY-1999;	99US-0135629P.	PR	02-AUG-1999;	99US-0146389P.
PR	24-MAY-1999;	99US-0136041P.	PR	03-AUG-1999;	99US-0147038P.
PR	25-MAY-1999;	99US-0136392P.	PR	04-AUG-1999;	99US-0147204P.
PR	27-MAY-1999;	99US-0136782P.	PR	04-AUG-1999;	99US-0147302P.
PR	28-MAY-1999;	99US-0137222P.	PR	05-AUG-1999;	99US-0147192P.
PR	01-JUN-1999;	99US-0137528P.	PR	05-AUG-1999;	99US-0147260P.
PR	03-JUN-1999;	99US-0137528P.	PR	06-AUG-1999;	99US-0147303P.
PR	04-JUN-1999;	99US-0137502P.	PR	06-AUG-1999;	99US-0147416P.
PR	07-JUN-1999;	99US-0137724P.	PR	08-AUG-1999;	99US-0147493P.
PR	08-JUN-1999;	99US-0138094P.	PR	08-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0148171P.
PR	10-JUN-1999;	99US-0138547P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148584P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	16-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	17-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	23-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	25-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	21-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140353P.	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155513P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157177P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.

PR	13-OCT-1999;	99US-0159295P.	XX	25-FEB-1999;	99US-0121825P.
PR	14-OCT-1999;	99US-0159329P.	PR	05-MAR-1999;	99US-0123180P.
PR	14-OCT-1999;	99US-0159330P.	PR	09-MAR-1999;	99US-0123548P.
PR	14-OCT-1999;	99US-0159331P.	PR	23-MAR-1999;	99US-0125788P.
PR	14-OCT-1999;	99US-0159337P.	PR	25-MAR-1999;	99US-0126264P.
PR	14-OCT-1999;	99US-0159338P.	PR	29-MAR-1999;	99US-0126785P.
PR	18-OCT-1999;	99US-0159584P.	PR	01-APR-1999;	99US-0127462P.
PR	21-OCT-1999;	99US-0160741P.	PR	06-APR-1999;	99US-0128234P.
PR	21-OCT-1999;	99US-0160767P.	PR	18-APR-1999;	99US-0128714P.
PR	21-OCT-1999;	99US-0160770P.	PR	19-APR-1999;	99US-0129845P.
PR	21-OCT-1999;	99US-0160814P.	PR	21-APR-1999;	99US-0130077P.
PR	21-OCT-1999;	99US-0160815P.	PR	23-APR-1999;	99US-0130449P.
PR	22-OCT-1999;	99US-0160980P.	PR	23-APR-1999;	99US-0130510P.
PR	22-OCT-1999;	99US-0160981P.	PR	28-APR-1999;	99US-0130891P.
PR	25-OCT-1999;	99US-0161404P.	PR	30-APR-1999;	99US-0131449P.
PR	25-OCT-1999;	99US-0161405P.	PR	30-APR-1999;	99US-0132048P.
PR	25-OCT-1999;	99US-0161406P.	PR	04-MAY-1999;	99US-0132407P.
PR	26-OCT-1999;	99US-0161359P.	PR	05-MAY-1999;	99US-0132484P.
PR	26-OCT-1999;	99US-0161360P.	PR	06-MAY-1999;	99US-0132485P.
PR	26-OCT-1999;	99US-0161361P.	PR	06-MAY-1999;	99US-0132486P.
PR	28-OCT-1999;	99US-0161920P.	PR	07-MAY-1999;	99US-0132487P.
PR	28-OCT-1999;	99US-0161922P.	PR	11-MAY-1999;	99US-0132863P.
PR	28-OCT-1999;	99US-0161933P.	PR	14-MAY-1999;	99US-0134218P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-MAY-1999;	99US-0134219P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-MAY-1999;	99US-0134221P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-MAY-1999;	99US-0134370P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-MAY-1999;	99US-0134768P.
PR	29-OCT-1999;	99US-0162142P.	PR	19-MAY-1999;	99US-0134941P.
PR	29-OCT-1999;	99US-0162142P.	PR	20-MAY-1999;	99US-0135124P.
PR	29-OCT-1999;	99US-0162142P.	PR	21-MAY-1999;	99US-0135353P.
PR	29-OCT-1999;	99US-0162142P.	PR	24-MAY-1999;	99US-0135829P.
PR	29-OCT-1999;	99US-0162142P.	PR	25-MAY-1999;	99US-0136021P.
PR	29-OCT-1999;	99US-0162142P.	PR	27-MAY-1999;	99US-0136392P.
PR	29-OCT-1999;	99US-0162142P.	PR	28-MAY-1999;	99US-0136782P.
PR	29-OCT-1999;	99US-0162142P.	PR	01-JUN-1999;	99US-0137222P.
PR	29-OCT-1999;	99US-0162142P.	PR	03-JUN-1999;	99US-0137528P.
PR	29-OCT-1999;	99US-0162142P.	PR	04-JUN-1999;	99US-0137502P.
PR	29-OCT-1999;	99US-0162142P.	PR	07-JUN-1999;	99US-0137724P.
PR	29-OCT-1999;	99US-0162142P.	PR	08-JUN-1999;	99US-0138094P.
PR	29-OCT-1999;	99US-0162142P.	PR	10-JUN-1999;	99US-0138540P.
PR	29-OCT-1999;	99US-0162142P.	PR	10-JUN-1999;	99US-0138847P.
PR	29-OCT-1999;	99US-0162142P.	PR	14-JUN-1999;	99US-0139119P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999;	99US-0139452P.
PR	29-OCT-1999;	99US-0162142P.	PR	16-JUN-1999;	99US-0139453P.
PR	29-OCT-1999;	99US-0162142P.	PR	17-JUN-1999;	99US-0139492P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139454P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139455P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139456P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139457P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139458P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139459P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139460P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139461P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139462P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139463P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139750P.
PR	29-OCT-1999;	99US-0162142P.	PR	18-JUN-1999;	99US-0139763P.
PR	29-OCT-1999;	99US-0162142P.	PR	21-JUN-1999;	99US-0139817P.
PR	29-OCT-1999;	99US-0162142P.	PR	22-JUN-1999;	99US-0139899P.
PR	29-OCT-1999;	99US-0162142P.	PR	23-JUN-1999;	99US-0140353P.
PR	29-OCT-1999;	99US-0162142P.	PR	23-JUN-1999;	99US-0140354P.
PR	29-OCT-1999;	99US-0162142P.	PR	24-JUN-1999;	99US-0140695P.
PR	29-OCT-1999;	99US-0162142P.	PR	28-JUN-1999;	99US-0140823P.
PR	29-OCT-1999;	99US-0162142P.	PR	29-JUN-1999;	99US-0140991P.
PR	29-OCT-1999;	99US-0162142P.	PR	30-JUN-1999;	99US-0141287P.
PR	29-OCT-1999;	99US-0162142P.	PR	01-JUL-1999;	99US-0141842P.
PR	29-OCT-1999;	99US-0162142P.	PR	01-JUL-1999;	99US-0142545P.
PR	29-OCT-1999;	99US-0162142P.	PR	02-JUL-1999;	99US-0142055P.
PR	29-OCT-1999;	99US-0162142P.	PR	06-JUL-1999;	99US-0142390P.
PR	29-OCT-1999;	99US-0162142P.	PR	08-JUL-1999;	99US-0142803P.
PR	29-OCT-1999;	99US-0162142P.	PR	09-JUL-1999;	99US-0142920P.

Query Match 83.3%; Score 1238; DB 3; Length 426;

Best Local Similarity 79.1%; Pred. No. 2.4e-123;

Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

QY 1 AKIIDGKLVAKQIREIAEIAKMDAIGVPEGLAVILVGRKDSQTYVRNKKKACEAVG 60

DB 135 AVIDGKAVAKLRDITTEVSMSKESIGVIGLAVILVGRKDSATYVRNKKKACDSVG 194

QY 61 IKSYEVNLPDSSEDEVKHIATFNSDPSVHGILVQLPLPHHMDNENILNAVSIKDVVG 120

DB 195 IKSEFEVRLAEDSEEVLSKVSFNDPSVHGILVQLPLPHHMDNENILNAVSIKDVVG 254

QY 121 FHEPLNIGRLAMGDRPFVPCPKGCMELLHRYGVEIKGRAVITGRSNIVGMPAALLIQ 180

DB 255 FHEPLNIGRLAMGDRPFVPCPKGCMELLHRYGVEIKGRAVITGRSNIVGMPAALLIQ 314

QY 181 KANATSVIHSNTKPEEITRQADIIVAAVGVANLVGRGSKWIKPGAALIDVGINPVDDPES 240

DB 315 REDATSVIHSNTKPEEITRQADIIVAAVGVANLVGRGSKWIKPGAALIDVGINPVDDPES 374

QY 241 PRGYRLVGDVCEYEASKIAGLITPVGGVPMPTIAMLNTLESKRIHKFK 292

DB 375 ARGYRLVGDVCEYEASKIAGLITPVGGVPMPTIAMLNTLESKRIHKFK 426

RESULT 4

AAG22173

ID AAG22173 standard; protein; 299 AA.

XX AC AAG22173;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 25001.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157533P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144085P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0158293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0158294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	21-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161359P.
PR	27-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0147260P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0147303P.	PR	26-OCT-1999;	99US-0161362P.
PR	06-AUG-1999;	99US-0147303P.	PR	28-OCT-1999;	99US-0161920P.
PR	06-AUG-1999;	99US-0147416P.	PR	28-OCT-1999;	99US-0161922P.
PR	09-AUG-1999;	99US-0147493P.	PR	28-OCT-1999;	99US-0161933P.
PR	09-AUG-1999;	99US-0147535P.	PR	29-OCT-1999;	99US-0162142P.
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	20-AUG-1999;	99US-0149929P.			
PR	23-AUG-1999;	99US-0149902P.			
PR	23-AUG-1999;	99US-0149930P.			
PR	25-AUG-1999;	99US-0150566P.			
PR	26-AUG-1999;	99US-0150884P.			
PR	27-AUG-1999;	99US-0151065P.			
PR	27-AUG-1999;	99US-0151066P.			
PR	30-AUG-1999;	99US-0151080P.			
PR	31-AUG-1999;	99US-0151303P.			
PR	01-SEP-1999;	99US-0151438P.			
PR	01-SEP-1999;	99US-0151930P.			
PR	07-SEP-1999;	99US-0152363P.			
PR	10-SEP-1999;	99US-0153070P.			
PR	13-SEP-1999;	99US-0153758P.			
PR	15-SEP-1999;	99US-0154018P.			
PR	16-SEP-1999;	99US-0154039P.			
PR	20-SEP-1999;	99US-0154779P.			
PR	22-SEP-1999;	99US-0155139P.			
PR	23-SEP-1999;	99US-0155486P.			
PR	24-SEP-1999;	99US-0155659P.			
PR	28-SEP-1999;	99US-0156458P.			
Query Match 69.7%; Score 1035; DB 3; Length 299;					
Best Local Similarity 67.4%; Pred. No. 7.9e-102;					
Matches 194; Conservative 45; Mismatches 49; Indels 0; Gaps 0;					
Qy	1	AKIIDGKLVAKQIREIAVEIAEKMDAIGVWPGVLAVILVGRKDSQTYVRNKKKCAEAVG	60		
Db	9	AKIIDGRAIAHTIRSEIAEEVRLSEKHKVPLGLAVIVGRKDSQTYVNRKKAEEVG	68		
Qy	61	IKSYEVNLPEDSSEDEVLKHIAIFNSDPVHGILVOLPLPHMNDENILNAVSIKDVVG	120		
Db	69	IKSFDVGLPEEVSEADLSIKVHELNSNPVHGILVOLPLPKHINEENILCAISIDKDVVG	128		
Qy	121	FHPLINTGLAMQSDPFFVPCTEKGCWELLHRVGVKRAVIGRSNIVGMPAALLQ	180		
Db	129	FHPLNTGKLAMKGRBFLFPCPTKGCLELLARSGVKIKQRAVVGRSVIGLPLVSLLL	188		
Qy	181	KANATYSIVHSNTKKPEEITRQADIVIAAVGVANLVGRSGWIKFGAAIIDVGINPDDPES	240		
Db	189	KADATVTVHSHTKDPFAIIRADIVIAACGQAHMIKGNWIKFGAAVIDVGTNAVSDPSK	248		
Qy	241	PRGYRLVGDVVCYEAEASKIAGLITPVGGVGPMTIAMLNTLESAKRI	288		
Db	249	XSGYRLVGDVDFAEASKVAGFITPVPGVGPMTVAMLLRNTVDGAKRV	296		
RESULT 5					
AAG08973					
ID AAG08973 standard; protein; 299 AA.					
XX AAG08973;					
XX AC					
XX DT 17-OCT-2000 (first entry)					
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6721.					
XX KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter;					

KW termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
PN 06-SEP-2000.
XX
PD
XX
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130591P.
XX 30-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 04-MAY-1999; 99US-0132407P.
XX 05-MAY-1999; 99US-0132484P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 18-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134370P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.
XX 10-JUN-1999; 99US-0138540P.
XX 10-JUN-1999; 99US-0138847P.
XX 14-JUN-1999; 99US-0139119P.
XX 16-JUN-1999; 99US-0139452P.
XX 16-JUN-1999; 99US-0139453P.
XX 17-JUN-1999; 99US-0139492P.
XX 18-JUN-1999; 99US-0139454P.
XX 18-JUN-1999; 99US-0139455P.
XX 18-JUN-1999; 99US-0139456P.
XX 18-JUN-1999; 99US-0139457P.
XX 18-JUN-1999; 99US-0139458P.
XX 18-JUN-1999; 99US-0139459P.
XX 18-JUN-1999; 99US-0139460P.
XX 18-JUN-1999; 99US-0139461P.
XX 18-JUN-1999; 99US-0139462P.
XX 18-JUN-1999; 99US-0139463P.
XX 18-JUN-1999; 99US-0139750P.
XX 18-JUN-1999; 99US-0139763P.
XX 21-JUN-1999; 99US-0139817P.
XX 22-JUN-1999; 99US-0139899P.
XX 23-JUN-1999; 99US-0140353P.
XX 23-JUN-1999; 99US-0140354P.
XX 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145513P.
PR 27-JUL-1999; 99US-0145518P.
PR 28-JUL-1999; 99US-0145519P.
PR 28-JUL-1999; 99US-0145591P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.


```
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140931P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144832P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.

PR 25-AUG-1999; 99US-0150556P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157855P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 68.4%; Score 1017; DB 3; Length 350;
Best Local Similarity 67.5%; Pred. No. 8.4e-100;
Matches 193; Conservative 36; Mismatches 57; Indels 0; Gaps 0;
Qy 3 IIDGKLVAKQIREIAVEIAKMKDAIGVVPGLAVILVGRKDSQTYVRNKKKACAVGK 62
Db 63 VIDGNVIAEEIRTKIISEVGRMKKAVGKVPGLAVILVGEQDSQTYVRNKKIACETGK 122
Qy 63 SYEVLNPEDSSSEDEVLKHATFNSDPSVHGILVOLPLPHHMDENILNAVSTEKVDGPH 122
Db 123 SVLAELPEDCTEGQIISVLRKNEDTSIHGILVOLPLPHHMDENILNAVSTEKVDGPH 182
Qy 123 PLNIGRLAQGRDPPFPVCTPKGOMELLHRYGVEIKGKRAVIGRSNIVGMPAALLQKA 182
Db 183 PLNVGNLAVRGREPLFVSCPTKGCVELLIRTVGVEIAGKNAVIGRSNIVGLPMSLLQKH 242
Qy 183 NATVSIVHSTNKKFEITRQADIVIAAGVANLVGRSNTKPGAAIIVDGINVDDPESPR 242
Db 243 DATVSIVHFTKDPHEITRKADIVIAAGIPNLVGRSMLKPGAVVIDVGTTPVDDSSCEP 302
```

Qy 243 GYRLVGDVCEEASKIAGLITPVGYPGWTIAMLLNTLESAKRI 288
Db 303 GYRLVGDVCEEALGVASALTPVGYPGWTITWLLCNTLEAKRI 348

RESULT 7

AAG36761

ID AAG36761 standard; protein; 352 AA.

AC AAG36761;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45097.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0128845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 04-MAY-1999; 99US-0132048P.

PR 05-MAY-1999; 99US-0132407P.

PR 06-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0132556P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134914P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 27-MAY-1999; 99US-0136021P.

PR 28-MAY-1999; 99US-0136392P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137228P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145152P.
PR 23-JUL-1999; 99US-014518P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147433P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.

PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 21-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 23-JUN-1999; 99US-0140698P.
PR 24-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 29-JUN-1999; 99US-0141287P.
PR 30-JUN-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 01-JUL-1999; 99US-0142058P.
PR 02-JUL-1999; 99US-0142390P.
PR 06-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142920P.
PR 09-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144008P.
PR 16-JUL-1999; 99US-0144088P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 23-JUL-1999; 99US-0145276P.
PR 26-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.

PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147418P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 12-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 27-AUG-1999; 99US-0151303P.
PR 30-AUG-1999; 99US-0151438P.
PR 31-AUG-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 22-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159233P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

Query Match

68.4%; Score 1017; DB 3; Length 498;

Best Local Similarity		67.5%;	Pred. No. 1.4e-99;
Matches 193;		Conservative 36;	Mismatches 57;
		Indels 0;	Gaps 0;
QY	3	IIDKLLVAKOIREBIAVEIIAKWKDAIGVPGCLAVILVGRSKDSOTYVRNKKKACEAVGIK	62
Db	211	VIDGNVIAEIRTKIIEVGKWKAKVGVPLGAVLVGEORDSOTYVRNKKACEETGIK	270
QY	63	SYENVLPEDSSEDEVLKHIAFTNSDDPSVHGILVOLPLPHMNDENILNAVSIKDVDFGH	122
Db	271	SVLAELPEDCTEGGIIISVLAKFNEEDTSIHGILVOLPLPQHLSKILNMVRLKDVDFGH	330
QY	123	PLNIGRLAMOGRRDPFPVCTPKGCMELLHRYGVYIKGRAVIVGRSNVGMPAALLLOKA	182
Db	331	PLNVGNLAMGRFLFPVSCTPKGCVEILLIIRTGVEIAGKNVAVTGRSNIVGLPMSLLQRH	390
QY	183	NATVSVIHSNTRKPEEITROADIVIAAVGVANLVGRSWIKPGAAIIDVGINPDDPSPR	242
Db	391	DATVSVTHAFTKDPKHITRKADIVIAAGIPNLVGRSLKFGAVVIDVGTTPVEDSSCEF	450
QY	243	GVELVGDVCEEASKTAGLITPVPGVGPMTIAMLSNTLESAKRI	288
Db	451	GVELVGDVCEEALGVASAITPVPGVGPMTIITMLCNTLEAAKRI	496
RESULT 9			
AAG14693			
ID	AAG14693 standard; protein; 350 AA.		
XX			
AC	AAG14693;		
XX			
DT	17-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 14652.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
PP	06-SEP-2000.		
PF	25-FEB-2000; 2000EP-00301439.		
XX			
PR	25-FEB-1999; 99US-0121825P.		
PR	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-MAR-1999; 99US-0125788P.		
PR	25-MAR-1999; 99US-0126264P.		
PR	29-MAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	19-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130449P.		
PR	23-APR-1999; 99US-0130510P.		
PR	28-APR-1999; 99US-0130891P.		
PR	28-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	30-APR-1999; 99US-0132407P.		
PR	04-MAY-1999; 99US-0132484P.		
PR	05-MAY-1999; 99US-0132485P.		
PR	06-MAY-1999; 99US-0132486P.		
PR	07-MAY-1999; 99US-0132487P.		
PR	07-MAY-1999; 99US-0132863P.		
PR	11-MAY-1999; 99US-0134256P.		
PR	14-MAY-1999; 99US-0134218P.		
PR	14-MAY-1999; 99US-0134219P.		
PR	14-MAY-1999; 99US-0134221P.		
PR	14-MAY-1999; 99US-0134370P.		

PR 05-MAY-1999; 99US-0132485P. 22-JUL-1999; 99US-0145087P.
PR 06-MAY-1999; 99US-0132486P. 22-JUL-1999; 99US-0145089P.
PR 06-MAY-1999; 99US-0132487P. 22-JUL-1999; 99US-0145102P.
PR 07-MAY-1999; 99US-0132488P. 23-JUL-1999; 99US-0145115P.
PR 11-MAY-1999; 99US-0132489P. 23-JUL-1999; 99US-0145145P.
PR 14-MAY-1999; 99US-0132490P. 23-JUL-1999; 99US-0145218P.
PR 14-MAY-1999; 99US-0132491P. 23-JUL-1999; 99US-0145224P.
PR 14-MAY-1999; 99US-0132492P. 26-JUL-1999; 99US-0145276P.
PR 14-MAY-1999; 99US-0132493P. 27-JUL-1999; 99US-0145913P.
PR 14-MAY-1999; 99US-0132494P. 27-JUL-1999; 99US-0145918P.
PR 18-MAY-1999; 99US-0132495P. 27-JUL-1999; 99US-0145919P.
PR 19-MAY-1999; 99US-0132496P. 28-JUL-1999; 99US-0145951P.
PR 20-MAY-1999; 99US-0132497P. 28-JUL-1999; 99US-0146386P.
PR 21-MAY-1999; 99US-0132498P. 02-AUG-1999; 99US-0146388P.
PR 24-MAY-1999; 99US-0132499P. 02-AUG-1999; 99US-0146389P.
PR 25-MAY-1999; 99US-0132500P. 03-AUG-1999; 99US-0147038P.
PR 27-MAY-1999; 99US-0132501P. 04-AUG-1999; 99US-0147204P.
PR 28-MAY-1999; 99US-0132502P. 04-AUG-1999; 99US-0147302P.
PR 01-JUN-1999; 99US-0132503P. 05-AUG-1999; 99US-0147192P.
PR 03-JUN-1999; 99US-0132504P. 05-AUG-1999; 99US-0147260P.
PR 04-JUN-1999; 99US-0132505P. 06-AUG-1999; 99US-0147303P.
PR 07-JUN-1999; 99US-0132506P. 06-AUG-1999; 99US-0147416P.
PR 08-JUN-1999; 99US-0132507P. 09-AUG-1999; 99US-0147493P.
PR 10-JUN-1999; 99US-0132508P. 09-AUG-1999; 99US-0147935P.
PR 10-JUN-1999; 99US-0132509P. 11-AUG-1999; 99US-0148171P.
PR 14-JUN-1999; 99US-0132510P. 12-AUG-1999; 99US-0148319P.
PR 16-JUN-1999; 99US-0132511P. 12-AUG-1999; 99US-0148341P.
PR 16-JUN-1999; 99US-0132512P. 13-AUG-1999; 99US-0148545P.
PR 17-JUN-1999; 99US-0132513P. 13-AUG-1999; 99US-0148684P.
PR 18-JUN-1999; 99US-0132514P. 16-AUG-1999; 99US-0149368P.
PR 18-JUN-1999; 99US-0132515P. 17-AUG-1999; 99US-0149175P.
PR 18-JUN-1999; 99US-0132516P. 18-AUG-1999; 99US-0149426P.
PR 18-JUN-1999; 99US-0132517P. 20-AUG-1999; 99US-0149722P.
PR 18-JUN-1999; 99US-0132518P. 20-AUG-1999; 99US-0149723P.
PR 18-JUN-1999; 99US-0132519P. 20-AUG-1999; 99US-0149929P.
PR 18-JUN-1999; 99US-0132520P. 23-AUG-1999; 99US-0149902P.
PR 18-JUN-1999; 99US-0132521P. 23-AUG-1999; 99US-0149930P.
PR 18-JUN-1999; 99US-0132522P. 25-AUG-1999; 99US-0150566P.
PR 18-JUN-1999; 99US-0132523P. 26-AUG-1999; 99US-0150884P.
PR 18-JUN-1999; 99US-0132524P. 27-AUG-1999; 99US-0151065P.
PR 18-JUN-1999; 99US-0132525P. 27-AUG-1999; 99US-0151066P.
PR 22-JUN-1999; 99US-0132526P. 27-AUG-1999; 99US-0151080P.
PR 22-JUN-1999; 99US-0132527P. 30-AUG-1999; 99US-0151030P.
PR 23-JUN-1999; 99US-0132528P. 31-AUG-1999; 99US-0151438P.
PR 23-JUN-1999; 99US-0132529P. 01-SEP-1999; 99US-0151930P.
PR 24-JUN-1999; 99US-0132530P. 07-SEP-1999; 99US-0152363P.
PR 28-JUN-1999; 99US-0132531P. 10-SEP-1999; 99US-0153070P.
PR 29-JUN-1999; 99US-0132532P. 13-SEP-1999; 99US-0153758P.
PR 30-JUN-1999; 99US-0132533P. 15-SEP-1999; 99US-0154018P.
PR 01-JUL-1999; 99US-0132534P. 16-SEP-1999; 99US-0154039P.
PR 01-JUL-1999; 99US-0132535P. 20-SEP-1999; 99US-0154779P.
PR 02-JUL-1999; 99US-0132536P. 22-SEP-1999; 99US-0155139P.
PR 06-JUL-1999; 99US-0132537P. 23-SEP-1999; 99US-0155486P.
PR 08-JUL-1999; 99US-0132538P. 24-SEP-1999; 99US-0155659P.
PR 09-JUL-1999; 99US-0132539P. 28-SEP-1999; 99US-0156458P.
PR 12-JUL-1999; 99US-0132540P. 29-SEP-1999; 99US-0156596P.
PR 13-JUL-1999; 99US-0132541P. 04-OCT-1999; 99US-0157117P.
PR 14-JUL-1999; 99US-0132542P. 05-OCT-1999; 99US-0157753P.
PR 15-JUL-1999; 99US-0132543P. 06-OCT-1999; 99US-0157865P.
PR 16-JUL-1999; 99US-0132544P. 07-OCT-1999; 99US-0158029P.
PR 16-JUL-1999; 99US-0132545P. 08-OCT-1999; 99US-0158232P.
PR 19-JUL-1999; 99US-0132546P. 12-OCT-1999; 99US-0158369P.
PR 19-JUL-1999; 99US-0132547P. 13-OCT-1999; 99US-0159293P.
PR 19-JUL-1999; 99US-0132548P. 13-OCT-1999; 99US-0159294P.
PR 19-JUL-1999; 99US-0132549P. 14-OCT-1999; 99US-0159329P.
PR 19-JUL-1999; 99US-0132550P. 14-OCT-1999; 99US-0159330P.
PR 20-JUL-1999; 99US-0132551P. 14-OCT-1999; 99US-0159331P.
PR 20-JUL-1999; 99US-0132552P. 14-OCT-1999; 99US-0159637P.
PR 20-JUL-1999; 99US-0132553P. 14-OCT-1999; 99US-0159638P.
PR 21-JUL-1999; 99US-0132554P. 18-OCT-1999; 99US-0159584P.
PR 21-JUL-1999; 99US-0132555P. 21-OCT-1999; 99US-0160741P.
PR 21-JUL-1999; 99US-0132556P. 21-OCT-1999; 99US-0160767P.
PR 22-JUL-1999; 99US-0132557P. 21-OCT-1999; 99US-0160768P.

PR	16-APR-1999;	99US-0129845P
PR	19-APR-1999;	99US-0130077P
PR	21-APR-1999;	99US-0130044P
PR	23-APR-1999;	99US-0130510P
PR	23-APR-1999;	99US-0130891P
PR	28-APR-1999;	99US-0131449P
PR	30-APR-1999;	99US-0132048P
PR	30-APR-1999;	99US-0132407P
PR	04-MAY-1999;	99US-0132484P
PR	05-MAY-1999;	99US-0132485P
PR	06-MAY-1999;	99US-0132486P
PR	06-MAY-1999;	99US-0132487P
PR	11-MAY-1999;	99US-0132863P
PR	11-MAY-1999;	99US-0134256P
PR	14-MAY-1999;	99US-0134218P
PR	14-MAY-1999;	99US-0134219P
PR	14-MAY-1999;	99US-0134221P
PR	14-MAY-1999;	99US-0134370P
PR	18-MAY-1999;	99US-0134768P
PR	19-MAY-1999;	99US-0134941P
PR	20-MAY-1999;	99US-0135124P
PR	21-MAY-1999;	99US-0135353P
PR	24-MAY-1999;	99US-0135629P
PR	25-MAY-1999;	99US-0136021P
PR	27-MAY-1999;	99US-0136392P
PR	28-MAY-1999;	99US-0136782P
PR	01-JUN-1999;	99US-0137222P
PR	03-JUN-1999;	99US-0137528P
PR	04-JUN-1999;	99US-0137502P
PR	07-JUN-1999;	99US-0137702P
PR	08-JUN-1999;	99US-0137809P
PR	10-JUN-1999;	99US-0138540P
PR	10-JUN-1999;	99US-0138847P
PR	14-JUN-1999;	99US-0139119P
PR	16-JUN-1999;	99US-0139452P
PR	16-JUN-1999;	99US-0139453P
PR	17-JUN-1999;	99US-0139492P
PR	18-JUN-1999;	99US-0139450P
PR	18-JUN-1999;	99US-0139455P
PR	18-JUN-1999;	99US-0139456P
PR	18-JUN-1999;	99US-0139457P
PR	18-JUN-1999;	99US-0139458P
PR	18-JUN-1999;	99US-0139459P
PR	18-JUN-1999;	99US-0139460P
PR	18-JUN-1999;	99US-0139461P
PR	18-JUN-1999;	99US-0139462P
PR	18-JUN-1999;	99US-0139463P
PR	18-JUN-1999;	99US-0139750P
PR	18-JUN-1999;	99US-0139763P
PR	21-JUN-1999;	99US-0139817P
PR	22-JUN-1999;	99US-0139899P
PR	23-JUN-1999;	99US-0140353P
PR	23-JUN-1999;	99US-0140354P
PR	24-JUN-1999;	99US-0140695P
PR	28-JUN-1999;	99US-0140823P
PR	28-JUN-1999;	99US-0140991P
PR	30-JUN-1999;	99US-0141287P
PR	01-JUL-1999;	99US-0141842P
PR	01-JUL-1999;	99US-0142154P
PR	02-JUL-1999;	99US-0142055P
PR	06-JUL-1999;	99US-0142330P
PR	08-JUL-1999;	99US-0142803P
PR	09-JUL-1999;	99US-0142920P
PR	12-JUL-1999;	99US-0142977P
PR	13-JUL-1999;	99US-0143542P
PR	14-JUL-1999;	99US-0143624P
PR	15-JUL-1999;	99US-0144005P
PR	16-JUL-1999;	99US-0144085P
PR	16-JUL-1999;	99US-0144086P
PR	19-JUL-1999;	99US-0144325P
PR	19-JUL-1999;	99US-0144331P
PR	19-JUL-1999;	99US-0144332P
PR	19-JUL-1999;	99US-0144333P

XX DT 02-JUL-2002 (first entry)
 XX STREPTOCOCCUS polypeptide SEQ ID NO 270.
 XX DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX KW antinflammatory; infection; vaccine; meningitis; gene therapy.
 XX OS Streptococcus pyogenes.
 XX PN WO200234771-A2.
 XX PD 02-MAY-2002.
 XX XX 29-OCT-2001; 2001WO-GB004789.
 XX PF 27-OCT-2000; 2000GB-00026333.
 XX PR 24-NOV-2000; 2000GB-00028727.
 XX PR 07-MAR-2001; 2001GB-00005640.
 XX XX (CHIR-) CHIRON SPA.
 XX FA (GENO-) INST GENOMIC RES.
 XX XX Telford J, Massignani V, Margarit Y Rosi, Grandi G, Frazer C;
 XX PI Tettelin H;
 XX XX WPI: 2002-352536/38.
 XX DR N-PSDB; ABN66178.
 XX XX New Streptococcus protein for the treatment or prevention of infection or
 XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
 XX PT detecting a compound that binds to the protein.
 XX XX Claim 1; Page 3182; 4525pp; English.
 XX XX The invention relates to a protein (ABP25413-ABP30895) from group B
 XX CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 XX CC the specification. The proteins have antibacterial and antiinflammatory
 XX CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
 XX CC antibodies that bind (1) are used in the manufacture of medicaments for
 XX CC the treatment or prevention of infection or disease caused by
 XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 XX CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 XX CC biological sample. (1) is used to determine whether a compound binds to
 XX CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 XX CC used as a vaccine or diagnostic composition. The disease caused by
 XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 XX CC acid encoding (1) may be used to recombinantly produce (1) and may be
 XX CC used in gene therapy. Antibodies to (1) are used for affinity
 XX CC chromatography, immunoassays, and distinguishing/identifying
 XX CC Streptococcus proteins
 XX SQ Sequence 286 AA;
 Query Match 50.0%; Score 743.5; DB 5; Length 286;
 Best Local Similarity 51.4%; Pred. No. 1.2e-70;
 Matches 149; Conservative 55; Mismatches 77; Indels 9; Gaps 4;
 QY 2 KIIDGLVAKQIRIEETAVETAKMKDAIGVVPGLAVILVGRSKDSQTYVRNKKKCAEAVGI 61
 Db 5 ELIDGKALAKQKQELAAKVNKKKGIQVGLVILVGGDDPASQVYVRNKKRAALTGVGF 64
 QY 62 KSEVNLPSDESEDEVLKHIAFNPSDPVHGILVOLPLPHNMNDENILNAVSIKXDVDGF 121
 Db 65 KSETVRLSEFICQELIAVIERVNAQNTTGHILVOLPLPHNHNKKIILAIIDPKKXVDGF 124
 QY 122 HPLNIGRLAQMGDRPFVCTPKGCMELLHRYGVETKGRVAVIGRSNIVGNPAALLIQ 181
 Db 125 HPNNTGHL-WSGR-PLMVPTSGINMELLREYNVNLGEGHVIIGRSNIVGNPAALLIQ 182
 QY 182 ANATVSIHNTKKPEITRQADIVIAAGVANLVRGSMWIKPGAIIIDVGINPVDPPESP 241

Db 183 KNATVTLTHSRTRQLEEVCRCADVLIIVAIGQHFITKQYIKDGAIVIDVGNRDDNG--- 239
 QY 242 RGYRLVGDVCEAEASKIAGLITPVPVGGVGPMTIAMLLNTLESKR-IHK 290
 Db 240 ---KLIGDAFDEVAEVAAKITPVPVGGVGPMTIAMLLQTYQSALRSTHK 286
 RESULT 15
 ABR52676
 ID ABR52676 standard; protein; 946 AA.
 XX AC ABR52676;
 XX XX 20-JUN-2003 (first entry)
 XX DE Protein sequence #SEQ ID 217.
 XX KW Multiprotein complex; eukaryote; drug target; diagnosis.
 XX OS Saccharomyces cerevisiae.
 XX PN EP1258494-A1.
 XX PD 20-NOV-2002.
 XX PF 20-DEC-2001; 2001EP-00130253.
 XX PR 15-MAY-2001; 2001EP-00111774.
 XX XX (CELL-) CELLZOME AG.
 XX PI Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 XX PI Marzioch M, Schultz JD, Superti-Furga GD;
 XX XX WPI: 2003-250078/25.
 XX DR N-PSDB; ACC60718.
 XX XX New isolated protein complexes useful for diagnosing a disease or
 XX PT disorder, or as a target for an active agent of a pharmaceutical,
 XX PT preferably a drug target in the treatment or prevention of disease or
 XX PT disorder.
 XX PS Disclosure; SEQ ID NO 217; 17pp + Sequence Listing; English.
 XX CC The invention relates to multiprotein complexes from eukaryotes. Proteins
 XX CC of the invention and DNA sequences encoding them are given in records
 XX CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
 XX CC obtainable by using a protein as a bait and isolating the set of proteins
 XX CC which is attached thereto from cells. Such protein complexes may comprise
 XX CC up to 30 distinct proteins. Protein complexes of the invention are useful
 XX CC for diagnosing a disease or disorder, or as a target for an active agent
 XX CC of a pharmaceutical, preferably a drug target in the treatment or
 XX CC prevention of a disease or disorder. Note: The sequence data for this
 XX CC patent is not represented in the printed specification, but is based on
 XX CC sequence information supplied by the European Patent Office. The complete
 XX CC document is available on CD-ROM
 XX SQ Sequence 946 AA;
 Query Match 49.8%; Score 739.5; DB 6; Length 946;
 Best Local Similarity 50.5%; Pred. No. 2e-69;
 Matches 149; Conservative 45; Mismatches 92; Indels 9; Gaps 2;
 QY 2 KIIDGLVAKQIRIEETAVETAKMKDAI-GVVPGLAVILVGRSKDSQTYVRNKKKCAEAVG 60
 Db 4 QVLDGKACQAFQSNITANEIKSQGHVGFAPFALAIQVGNRPDSATYVRNKKKAAEEAG 63
 QY 61 KSEVNLPSDESEDEVLKHIAFNPSDPVHGILVOLPLPHNMNDENILNAVSIKXDVDG 120
 Db 64 IVANFIHLDESATEFEVLYRYVDQNLNEDPHTHGIIVQLPLPAHLDEDRITSRVLAEXDVG 123
 QY 121 FPLNIGRLAQMGDRPFVCTPKGCMELLHRYGVETKGRVAVIGRSNIVGNPAALLIQ 180

Db 124 FGPTNIGELNKNHGFPLPCTPKGIIELLHKANVTIEGSRVWIGRSDIVGSPVAELLK 183
QY 181 KANATVSIVHSNTKKPEETTROADIVIAAVGVANLVGSGWIKP-----GAAIIDVGI 232
Db 184 SLNSTVTITHSKTRDIASYLHDADIVWAIGQFEFVKGEWFKPRDGTSSDKKTWVIDVGT 243
QY 233 NPVDDPESPRGYRLYGDVCYEAEASKIAGLITPVGVGPMTIAMLLSNTLESAGR 287
Db 244 NYVADPSKSGFKCVGDVEFNEAIKVVHLITPVGVGPMTVAMLMQNTLIAAKR 298

Search completed: July 24, 2004, 06:36:16
Job time : 106 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 24, 2004, 06:41:26 ; Search time 5374 Seconds
(without alignments)
2355.074 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIIDGKLVAKRIIEAVE.....TIAMLLSNTLESAKRIHKFX 292

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2.1/USPFO_spool_p/US10723061/runat_21072004_090400_7108/app_query.fasta_1.455
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=500 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10723061@cgn2.1_7406 @runat_21072004_090400_7108 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: gb.pa.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1486	100.0	1116	6	AX653969	Sequence
2	1238	83.3	1083	6	AX505846	Sequence
3	1238	83.3	1083	6	AX651582	Sequence
4	1238	83.3	1132	8	BT000066	Arabidops
5	1238	83.3	1498	8	AY059944	Arabidops
C 6	1216.5	81.9	123454	8	AP005851	Oryza sat
C 7	1116.5	75.1	91849	8	F6M23	Arabidops
C 8	1116.5	75.1	197975	8	ATCHRIV2	Arabidops
C 9	1054.5	71.0	92675	8	AP006402	Lotus cor
10	1046.5	70.4	28160	8	AP006092	Lotus cor
11	1035	69.7	1182	8	AY085538	Arabidops
12	1034	69.6	1219	8	AF030516	Pisum sat
13	1034	69.6	1219	8	PSA011589	Pisum sat
14	1030	69.3	900	8	BT004612	Arabidops
15	1012	68.1	1118	8	AK103505	Oryza sat
16	997.5	67.1	1282	8	BT004325	Arabidops
17	988	66.5	1358	8	AK059603	Oryza sat
18	949	63.9	1269	8	AX654551	Sequence
19	892	60.0	127125	2	AC123899	Medicago
C 20	833	56.1	107541	8	AC005499	Arabidops
21	830.5	55.9	52211	8	AC069472	Arabidops
22	830.5	55.9	86536	8	AP002047	Arabidops
23	802	54.0	188080	2	AP005811	Oryza sat
24	801	53.9	114031	8	AC139354	Medicago
25	792	53.3	480	8	AF439725	Zea mays
26	765.5	51.5	4160	8	AF291429	Yarrowia
C 27	763.5	51.4	195269	6	AX417035	Sequence
28	763.5	51.4	33050	1	AL596168	Listeria
29	763.5	51.4	349980	6	AX417044	Sequence
30	755.5	50.8	250050	1	AL591978	Listeria
31	755.5	50.8	349980	6	AX416832	Sequence
32	751.5	50.6	5635	6	AX416832	Sequence
C 33	743.5	50.0	10029	1	AE010067	Streptoco
C 34	743.5	50.0	10165	1	AE006583	Streptoco
C 35	743.5	50.0	50463	1	AE014159	Streptoco
36	743.5	50.0	310950	1	AP005443	Streptoco
37	741	49.9	10197	1	AB009687	Brucella
C 38	741	49.9	11059	1	AE014572	Brucella
39	740.5	49.8	3116	5	BC045019	Xenopus l
40	739.5	49.8	2000	6	AX594564	Sequence
41	739.5	49.8	2000	6	AX818306	Sequence
42	739.5	49.8	2000	6	AX829336	Sequence
43	739.5	49.8	3529	8	SC1GR204W	S.cerevisia
44	739.5	49.8	4883	8	YSCADE3	Saccharomyc
45	739.5	49.8	18230	8	SCPEGR620	S.cerevisia

ALIGNMENTS

RESULT 1


```

QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGGATTCATGATGATTCCTTCTGTCATGGAATCCTTGTTCAGTTGCCCTTCGCCA 504
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGGATGAACAGAACATATCTGAATCGCGTTAGTAGAGAAAGATGTTGACGGA 564
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db 565 TTTTCATCGCTAAATATTTGAGCGCTTCCCATGCTGGAAGAACCTTATTGCTTCCT 624
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 625 TGTACTCCAAAGGATGATTCAGTGTGTGCATAGATACAACTTGAATCAAGAAAG 684
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGACGGCTGTATTCGGAAGAGATTAATGTCGTTATGCGGTATGCGCTCTTTTACTGCG 744
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 745 AGGAGGATGCAACCGTTAGCATTCATTCAGAACCAAGAACCTGGAAGAAATCACA 804
QY 201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaLeuLeuValArgGlySerTrp 220
Db 805 AGAAAGCTGATATTAATCTCAGCTGTTCGACGACCAACATGCTCAGAGAAAGCTGG 864
QY 221 IleLysProGlyAlaAlaIleAspValGlyLeuAsnProValAspAspProGluSer 240
Db 865 ATAAACCGCGCGACGCTCATCGATGTTGGATTAATCTCTGTGAGATCCAAAGTGCCT 924
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 925 CGCGCTGGATATTCGATGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAAGTTGCATCA 984
QY 261 LeuIleThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 985 GCCATCACCTGTCTCTGGCGGTGTAGGACCAATGACCATGACCATGCTTCTATCCAAC 1044
QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 3
AX651582 1083 bp DNA linear PAT 22-MAR-2003
LOCUS Sequence 393 from Patent WO03000898.
DEFINITION AX651582
ACCESSION AX651582
VERSION AX651582.1 GI:29154400
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Kacagiri,F., Qian,S., Tao,F., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 393 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1..1083
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"

ORIGIN
Alignment Scores: 2.01e-98 Length: 1083
Pred. No.: 1238.00 Matches: 231
Score: 91.44% Conservative: 36
Percent Similarity:

```

```

Best Local Similarity: 79.11% Mismatches: 25
Query Match: 83.31% Indels: 0
DB: 6 Gaps: 0
US-10-723-061-18 (1-292) x AX651582 (1-1083)
QY 1 AlaLysIleAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 205 GCATTTGTATAGATGGAAGCTGTGCAAAAAGATAGAGATGAATCAATTTGAA 264
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 265 GTTTCGAAGATGAAGAAATCAATTCGTTGTGATCTCTGCTTTAGCAGTAAATCCTCTGCT 324
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 325 GACAGAAAAGATCTGCAACTTATGTGAGAACAGAGAAAGCTTGTGACTCCGTTGGA 384
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 385 ATCAATTCGTTGCAAGTTCGCTAGCTGAAGATTCATCAGAAAGAGAGGTGTTGAAATCT 444
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 445 GTATCAGGATTCATATGATGATCCTTCTGTCATGGAATCCTTGTTCAGTTGCTCTGCCA 504
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 505 TCGCATATGGATGAACAGAACATATCTGAATGCGGTAGTATAGAGAAAGATGTTGACGGA 564
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db 565 TTTTCATCCCTAAATATTGACGCGCTTGCATGCGTGGAGAGAACCCCTTATTTCGTTCTCT 624
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 625 TGTACTCCAAAGGATGATTCAGTGTGTGCATAGATACAACTTGAATCAAGAAAG 684
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 685 AGACGGCTGTATTCGGAAGAGATTAATGTCGTTATGCGGTATGCGCTCTCTTTTACTGCG 744
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 745 AGGAGGATGCAACCGTTAGCATTCATTCAGAACCAAGAACCTGGAAGAAATCACA 804
QY 201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db 805 AGAAAGCTGATATTAATCTCAGCTGTTCGACGACCAACATGCTCAGAGAAAGCTGG 864
QY 221 IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 865 ATAAACCGCGCGAGCTCCTCATCGATGTTGGATTAATCTCTGTGAGATCCAAAGTGTCT 924
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 925 CGCGCTGGATATTCGATGTTGGAGACATTTGCTACGAGGAGGCTAGCAAAAGTTGCATCA 984
QY 261 LeuIleThrProValProGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 985 GCCATCACCTGTCTCTGGCGGTGTAGGACCAATGACCATGACCATGCTTCTATCCAAC 1044
QY 281 ThrLeuGluSerAlaLysArgIleHisLysPheLys 292
Db 1045 ACTTTAACATCAGCTAAGAGGATTCACAACTTCCAG 1080

RESULT 4
BT000066 1132 bp mRNA linear PLN 19-SEP-2002
LOCUS Arabidopsis thaliana Unknown protein (At4g0620) mRNA, complete cds.
DEFINITION BT000066
ACCESSION BT000066
VERSION BT000066.1 GI:23197715

```


Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE AUTHORS

Nguyen, M., Karlin-Neumann, G., Southwick, A., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Bath, J., Carninci, P., Chen, H., Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A., Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K., Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K., Shinzaki, K., Ecker, J., Theologis, A. and Davis, R.W.

Direct Submission
Submitted (22-OCT-2001) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: "RIKEN Arabidopsis Full-length cDNA": Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinzaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Nguyen, M., Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J., Bowser, L., Jones, T., Bath, J., Chen, H., Cheuk, R., Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Nguyen, M. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinzaki, K. (RIKEN GSC) and Davis, R.W. (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers
1..1498
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="IV"
/clones="RAPL05-08-M11"
/note="This clone is in pBluescript"

gene

1..1498
/gene="At4g00620; F6N23.26"

CDS

187..1269
/gene="At4g00620; F6N23.26"
/codon_start=1
/product="Unknown protein"
/protein_id="AAL24426.1"
/db_xref="GI:16649149"

/translation="MAGSMFTDCSGTTSRLHNRSGTFLRCVQGLRQTASG
RGCRSSSPISADTKSEGGAIVIDGKAVAKIRDEITIEVSRMKEISIGVPL
AVILGDKDSATVVRNKKACDSVGIKSEFVRLAEDSEBEVLKSVGFNDPSVHG
ILVQLPSPHMDQNLINAVSIEKVDGPHPLATGRMBREPLFVPCPKGCIELL
HRYNIEIKKRAVIGNSNIVGMPAALLQREDATVSIHSRTKNPEITREADILIS
AVGQPMVVRGSKVIGPAVLIDVGINPDEPSAARGYRLVGDICYEERASKVASAITFP
GGVGPMTIAMLLSNTLTSAKRIHNFQ"

ORIGIN

Alignment Scores:
Pred. No.: 2,97e-98 Length: 1498
Score: 1238.00 Matches: 231
Percent Similarity: 91.44% Conservative: 36
Best local Similarity: 79.11% Mismatches: 25
Query Match: 83.31% Indels: 0
DB: 8 Gaps: 0

US-10-723-061-18 (1-292) x AY059944 (1-1498)

QY 1 AlalysilleAspGlyLysLeuValAlalysGlnIleArgGluGluAlaValGlu 20
Db 391 GCAATGTGAATAGTGAAGGCTGTGGCAAAAAGATTAGAGATGAAATCACATTTGAA 450

QY 21 IleAlalysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 451 GTTTCAGGAATGAAGGAATCAATGGTGTGATTCCTGTTAGCAGTAATCTTGTGGT 510

QY 41 SerArgLysAspSerGlnThrTyValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 511 GACAGAAAAGATTCTGCAACTTAITGTAGGAAACAAGAAAGCTGTGTGACGCTTGA 570

QY 61 IleLysSerTyGluValAsnLeuProGluAspSerSerGluAspGluValleuLysHis 80
Db 571 ATCAATCGTTTGAAGTTCGTCTAGCTGAAGATTATCATCAGAAAGAGAGGTGTGAAATCT 630

QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 631 GTATCAGATTCAATGATGATCCCTCTCTCATGGAATCCCTGTTTCAGTTGCTCTGCCA 690

QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 691 TCCCATATGGATGAACAGAACATATCTGAATCGGTTAGTATAGAGAAAGATGTTGACCGA 750

QY 121 PheHisProLeuAsnIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 140
Db 751 TTTCATCCGCTTAATATTGGACGGCTTGCCATGCGTGGAGAGAACCTTATTCTGTTCT 810

QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyGlyValGluIleLysGlyLys 160
Db 811 TGTACTCCAAAAGGATGCATTGAGTTGTGCATAGATACACATTTGAAATCAAGGAAAG 870

QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 871 AGAGCGGTTGTTATCGGAAGGAGTAACATTGTCGTATGCCAGCTGCTCTTTTACTGCAG 930

QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluLeuThr 200
Db 931 AGGAGGATGCAACCGTTAGCATTTCCATTCAAGAACCAAGAACCTTGAAGAAATCACA 990

QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db 991 AGAGAAGCTGATATTATATCTCAGCTGTGTGACAGCCCAACATGCTCAGAGAGAGCTGG 1050

QY 221 IleLysProGlyValAlaIleAlaIleAspValGlyIleAsnProValAspAspProGluSer 240
Db 1051 ATAAACCCGGCGCGATCCTCATCGATGTGGATTAATCTCTGAGGATCCAAGTCT 1110

QY 241 ProArgGlyTyArgLeuValGlyAspValCysTyGluGluAlaSerLysIleAlaGly 260
Db 1111 GCGCGTGGATATCGATTGTTGGAGACATTTGCTACGAGAGGCTAGCAAAAGTTGCATCA 1170

QY 261 LeuIleThrProValProGlyValGlyValProMetThrIleAlaMetLeuLeuSerAsn 280
Db 1171 GCATCACACCTGTTCTCGCGGTGTAGGACCAATGACCATAGCCATGCTTCTATCCAAAC 1230

QY 281 ThrLeuGluSexAlaLysArgIleHisLysPheLys 292
Db 1231 ACTTTAATCATCAGCTAAGAGGATTTCACAACTTCCAG 1266

RESULT 6
AP005851/c 123454 bp DNA linear PLN 01-AUG-2003
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
DEFINITION BAC clone:OSJNB0088N06, complete sequence.
ACCESSION AP005851
VERSION AP005851.2 GI:33383227
KEYWORDS HIG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.

REFERENCE 1
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 2, BAC
clone:OSJNB0088N06
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 123454)
AUTHORS Sasaki, T., Matsumoto, T. and Katayose, Y.

TITLE Direct Submission
JOURNAL Submitted (24-OCT-2002) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kamondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsakaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT On Jul 31, 2003 this sequence version replaced gi:24371211.
 The orientation of the sequence is from M13rev to -21M13 of the BAC
 clone.

FEATURES Location/Qualifiers
 source 1..123454
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="OSJNB0088N06"

ORIGIN
 Alignment Scores:
 Pred. No.: 4,48e-94 Length: 123454
 Score: 1216.50 Matches: 292
 Percent Similarity: 37.87% Conservative: 0
 Best Local Similarity: 37.87% Mismatches: 0
 Query Match: 81.86% Indels: 479
 DB: 3 Gaps: 3

US-10-723-061-18 (1-292) x AP005851 (1-123454)

Qy	1	AlaLysLeileAspGlyLysLeuValAlaLysGlnIleArgGluGluLeuValGlu	20
Db	77773	GCCAAAGATCATTGATGGGAAGTGGTGGCAAGACAGATGAGAGGAATCGCTGTGAG	77714
Qy	21	IleAlaLysMetLysAspAlaIleGlyValProGlyLeuAlaValIleLeuValGly	40
Db	77713	ATCGCAAGATGAAGATGCAATTTGGGTTGTGCTGGCTGGCAGTCATCTAGTTGGG	77654
Qy	41	SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysLysLysLysLysLys	60
Db	77653	TCAGGAAGATTCTCAACGATGATGTCGCAACAGAGAGAGAGAGAGAGAGAGAGAG	77594
Qy	61	IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis	80
Db	77593	ATCAAGTCATATGAGTTAAATTTCCGGAAGACAGCTCTGAGCATGAGGTTCTCAAGCAC	77534
Qy	81	IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValCateuProLeuPro	100
Db	77533	ATCGCAACATTTAACAGTATCCGTCGGTGCATGGCACTTGGTTTCAGTTGCCCTACCT	77474
Qy	101	His-----	101
Db	77473	CATGTAAGTTTTCACAAATTTCTTTTGGGCCAAATGCTCTGTATAACCTATCCACTGATA	77414
Qy	101	-----	101
Db	77413	TTGAATTTGGCCCTTACCAACTGAACCTGGTAGTTAAATGTATTTTAGTGAGGTACATT	77354
Qy	101	-----	101
Db	77353	CCTAGATATGATGGCAGTATGGCACACCATGGAACATTACTTTATGCACAAATTTTC	77294
Qy	101	-----	101
Db	77293	TGGAATAATTTGATATTTGGTGCCATCTCCAAATTTGTACATAATGTTTCTATTATGT	77234
Qy	101	-----	101
Db	77233	TAACTGGTTTGTATGCTCTTTGTGAAGATATCTGGTCTTAGAAAAATGTCGAACA	77174
Qy	101	-----	101
Db	77173	GATCCTTTTGTAGTAGTCGTCGTCATTCCTTTGTGGTCTTCGAGTTTGTGTGAC	77114

Qy	101	-----	101
Db	77113	CATTGTTTTTAANTAGATGGTCTATACTGGTCTGCAGTTCTCTGTTAATCTTGTGC	77054
Qy	101	-----	101
Db	77053	ATTGTTTGTAGCATATTAATGATGGGATATTTTGAATAATATTTTGTGATCAAGATACTT	76994
Qy	101	-----	101
Db	76993	TGAATGATTTTACCTGTTAGCATTTAAAGAGGTATTTGCTGCTAGCATCACAAAGTTT	76934
Qy	102	-----HisMet	103
Db	76933	TCCATGCTTCCATCCACTGAGCATCTGATTTCTGCTCTCTGATGGTTTGCAGCATATG	76874
Qy	104	AsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGlyPheHisPro	123
Db	76873	AATGATCAGAACATTTTGAATGCTGTAGTATTGAGAAGCATTTGATGGATTTCATCCA	76814
Qy	124	LeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValProCysThrPro	143
Db	76813	CTGAACATTGGACGACTTGCATGCAAGTTCGGATCCGTTCTTTGTTCCATGCACCCCT	76754
Qy	144	LysGlyCysMetGluLeuLeuHisArgTyrGlyValGluLeuLysGlyLysArgAlaVal	163
Db	76753	AAAGGATGCATGGAATTTACTACACAGATATGGAGTTGAAATCAAGGAAGAGAGCTGT	76694
Qy	164	ValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln-----	180
Db	76693	GTAATTGGGGGAGCAATATTTGGGATGCTCTGCTGCAATTTACTGCAAGTATGAAT	76634
Qy	180	-----	180
Db	76633	AAATCTCACTTCAATTAGTTGGAGAAATTTTATTTGGAGTGTTTAATTAACATCTTTTAT	76574
Qy	181	-----LysAlaAsnAlaThrValSerIleValHisSerAsnThrLys	194
Db	76573	GCCATATGATTAAGTATGAGAAGCCCAACGCACTGTAGCATTTGATTTCAATCAATCAAG	76514
Qy	195	LysProGluGluIleThrArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsn	214
Db	76513	AAGCCTGAGCAATAACAAGACAAGCAGATATTTGTTATCGACCTGTGGAGTTGCTAAT	76454
Qy	215	LeuValArgGlySerThrIleLysProGlyAlaAlaIleLeuAspValGlyIleAsnPro	234
Db	76453	CTGGTCAGAGGAGTTGGATAAAGCCTGGTGTCTATTATTGATTTGGCATCAATCCA	76394
Qy	235	ValAsp-----	236
Db	76393	GTGTGATGTAAGACACCTACTTTTCTGATGGCTTCCTTAATCTTCTCCCTAACATCCA	76334
Qy	236	-----	236
Db	76333	ACAGATCTGTTGCTTTTAAATGCCAACCCATTTCTAGATGGATAGTTTATGTCATGAAA	76274
Qy	236	-----	236
Db	76273	ACTGGGCATGTCATAGCCAAAGTCCCTTATATTTTGGGAGCGAGGTAGTTTGAACCTGCT	76214
Qy	236	-----	236
Db	76213	AATAGATTTTCCAATGTTACTGGTCCCTGGAATATCACACATTTGGTGGGATCCAGATG	76154
Qy	236	-----	236
Db	76153	GCTAATGGTGTATTGTTGTTTAAATAAATAATATATATATGCTTTTGCACCTTCATGTTT	76094
Qy	236	-----	236
Db	76093	CTAATCATGTTTGGTTTCATATATGCAAAATGTCAATCTTAGAAGGAATATAGTTTTTT	76034
Qy	236	-----	236

St. Louis, MO 63108, USA
e-mail: rwillson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEFHBRING COSMID INFORMATION:

The 5' clone is F5110, 200 bp overlap; 3' clone is F15P23, 900 bp overlap. Actual start of this clone is at base position 104960 of CELF5110; actual end is at 91149 of CELF6N23

The clone sequenced to the left is F5110. The actual start of this clone is at base position 104960 of F5110; actual end is at base position 91149 of F6N23.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES		source
Location/Qualifiers		
1..91849	/organism="Arabidopsis thaliana"	
	/mol_type="genomic DNA"	
	/cultivar="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="IV"	
	/map="unknown"	
	/clone="F6N23"	
	complement(187..907)	
	/gene="F6N23..23"	
	complement(join(187..347,396..512,598..907))	
	/genes="F6N23..23"	
	/codon_start=1	
	/evidence=not experimental	
	/protein_id="AAC13624.2"	
	/db_xref="GI:6382044"	
	/translation="MENLSPENDDDHQSFRIDQDQSAVTPVYVSTMSIDSFVYPR TCSETTSGFDQIDETNSFCSEPCNWPVLTESKSKCLGLEQSNCLVQVEISE PELETWKERFALLIGEDMSGGKGVCTATVTSNATNLNDCKNDTSFNDVCSYSW TESEVAVNRRESIVEERNEISFCMLRHCRIHP"	
	complement(4919..7497)	
	/gene="F6N23..22"	
	complement(join(4919..4993,5078..5488,5567..6055, 6157..6213,6498..6594,6935..7222,7349..7497))	
	/genes="F6N23..22"	
	/codon_start=1	
	/evidence=not experimental	
	/protein_id="AAC13623.1"	
	/db_xref="GI:3047112"	
	/translation="MSLTWADGVAAAGRSKQNSLLRKQLALAVPSVQMSYAFWSS SLTPQGVLEMGECYDGMKKRKYSEHYKGLQSKELKYLKLSMGDSGTTVST THNQLDNDCHSTMSLSPDLSDEWYLVMSYVFSQCLPGRASATGETIWL CNAQYAKNLFSLRLATVCPFGVIEGLVTELI SEDHNLNKSCLMEISAH QDNDEKMEIKI SEKHOLPLGIDSDHLKRTI STVLNYSADRSQKDNKRHRQP NIYTSPEGSFLNKKCEQVSGVOKKSONVLRKILHDVPLMHKTRMPSPNSGLN QDDPSDRKNEKFSVLRVTPVNEVDKESILNNTIKYLQELAEARVEELSCWGSVN"	

Db	76033	TTTTCAAGAAAGGAACATTAATTGTCTAGATCCTCGAGCTGGATCGAAGCAGGT	75974
Qy	236	-----	236
Db	75973	AAGAACTACACAGTAGACCCTCTGTGATCCCACTGTGGACACCACTTTATATAAAC	75914
Qy	236	-----	236
Db	75913	TTAAATTCCTTATTAAATTTTGATGATCCCATATCTTTCTGGTAAATTTCTTCAATT	75854
Qy	236	-----	236
Db	75953	GTATATCTGTTGGCAATATGATGATTAATCAATTTAGTACAAAGCCTTCCTTAAT	75794
Qy	236	-----	236
Db	75793	GTTGGCTTGTCATATCTTAAAGTTTGTGTATGTTTGTATCTCACTCACTTAATA	75734
Qy	236	-----	236
Db	75733	CCTCAGAAAGTTTCCAAAGAAATTTGGTGACTTGGCTGCTTTCCCTACACCCCTTCT	75674
Qy	237	-----AspProGluSerPro	241
Db	75673	GATTTCTTGCATGCTATGCTTACTAGTATTTGAAATGTCTCAGGATCCAGAAAGCCT	75614
Qy	242	ArgGlyTyrArgLeuValGlyAspValCysTyGluGluAlaSerLysIleAlaGlyLeu	261
Db	75613	CGAGGTTATCGGCTGGTGGAGATGTGTCTACGAGGAGGCTCCAAAGATTGAGACTA	75554
Qy	262	IleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThr	281
Db	75553	ATCAGCCAGTCTCTGTGTGTGTCGGCCATGACAAATTCGATGCTTTTGTGGAACACA	75494
Qy	282	LeuGluSerAlaLysArgIleHisLysPheLys	292
Db	75493	CTTGAGTCAGCTAAAGGATACACAAATTCAAA	75461

RESULT 7					
F6N23/c					
LOCUS	F6N23	91849 bp	DNA	linear	PLN 12-NOV-1999
DEFINITION	Arabidopsis thaliana BAC F6N23.				
ACCESSION	AF058919				
VERSION	AF058919.2	GI:6382042			
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 91849)				
AUTHORS	Geisel,C.				
TITLE	The sequence of A. thaliana F6N23				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 91849)				
AUTHORS	Washington University Genome Sequencing Center.				
TITLE	The A. thaliana Genome Sequencing Project				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 91849)				
AUTHORS	Waterston,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-APR-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
REFERENCE	4 (bases 1 to 91849)				
AUTHORS	Waterston,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-NOV-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA				
COMMENT	On Nov 12, 1999 this sequence version replaced gi:3047100.				
	Submitted by:				
	Genome Sequencing Center				
	Department of Genetics, Washington University,				

```

FVERQRTTNLNDLSVLIETSGNYDDSTKIDNSGETEQVTVFRDXKTHLRVRLKETE
VWIEVCSXDYIVADIMETLSNLHMDAFSVRSHTLNKFLTLNKAIFRGAASVYGM
IKRELRRVIGDLF
10235..12675
/gene="F6N23.1"
join(10235..10495,10693..10896,10998..11192,11341..11505,
11633..11893,11998..12216,12315..12534,12617..12675)
/gene="F6N23.1"
/note="similar to the family of glycosyl hydrolases"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13634.1"
/db_xref="GI:3047123"
/translation="MAIRLNHSVIPSVKLGAPTRVARSRLSPFSGVDRGVSTFSGA
RPIVLAKVLRASSTEDRVIDDDDDSTDLQVSEFHLDDSLGLVDNNSVVEPEE
LLDQLATLRSVNVGVDCWGMGIVESHDFQVNMWSGYKKLFQMIRELKQIVVMWF
HECGNVDGVHIOIPEWVREIQSNPDIFTDSAGRNTGCLTWGIDKQRLRGRTA
LEYFYDMRSRFEVDFEPEEKIPIEIVGLGPCGELRYPSPAGFWMYPGIGFPOC
YDKYLMNSLKEAAEVGHSGFWGSGPIENTYSTPHGTGPRDGDYDYSYGRFTLM
YSRLVDHGRVRLAMANLAFEGTCIAAKLSGIHWMYKTAASHAEITAGFYNSNRDGY
GPTAAVEKHDAALNTCVLELRLDQHEDFPEALADPEGLVQVQVNLNAWDASIPVASE
NALPCDRGYNKILNAPLTDPDGRHLSCFTYILNPTLMESQNFKEFERFLKRMH
GEAPDGLGAPGQETNPE"
complement(13027..14791)
/gene="F6N23.2"
complement(join(13027..13335,13480..13615,13712..13840,
13943..14001,14087..14791))
/gene="F6N23.2"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13622.1"
/db_xref="GI:3047111"
/translation="WSILCCVPLECVYCLGCTHMLKWKLYSAGHSESNWGLATSD
FEPPIRCLILAVAEENLHDPWAPDGVGDIDPNHVLKDDYDTEGRVTPMYILD
HENGVDVLAIRGLNAKEDYAVLDNKLQKFGGIVHNGLLKAAWMEFEHVL
RELLEANPSYITFVHSGAGVSLVLFVIONVRVLGNIEKRKIFAFAPROMS
LHLAVYADVINSVLVHDFLPRITALENFKSICFLCLCLCTLDFTFEERL
KDARLYAPGRYIHLVVRKPLRGYPVVRVTAIPVDGRFQIVLSNATADHAIWI
ERESQALDKIVRQKSIIVDHDEYRAALMAKASLNIIPMSPPSYGTHTDEGEBSA
GSGMGSPSGWSFGKMRKWDQFIDCHFPVNDNSEHMFQKQESQALLCGQET"
15539..16687
/gene="F6N23.2"
join(15539..15554,15672..15764,15953..16026,16133..16208,
16307..16353,16466..16687)
/gene="F6N23.2"
/note="Similar to protein kinase; coded for by A. thaliana
cDNA H36947; coded for by A. thaliana cDNA H37158"
/codon_start=1
/protein_id="AAC13612.1"
/db_xref="GI:3047101"
/translation="WNTSEVVEFLGNVTLQRLPSSLSKRISSEVVVFKGYDRGVYVR
ENQYDGVYFLQGVQRAQVLSAGEANYQEFPLKRYDFGFGIFGDVYSDVAVATE
ITCLLMSDHRALLEIKSVSDSKERCLVEDILVLEPLDVRVSPNCFSSISVLMIK
VLTIFHYHSHSESEL"
16976..19151
/gene="F6N23.3"
join(16976..17111,17262..17415,17785..17889,18316..18417,
18618..18666,19056..19151)
/gene="F6N23.3"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13635.1"
/db_xref="GI:3047124"
/translation="NEIYSNCHDVTIVASQALAAASKTVEMKLVHNFHCFYLLVGTI
NIPYIVNRDGNRPNATRSVDARQKRTIFLPASQVSCYIPEIFHYESLDYRQ
VATEISVFPIDIREPNRSTKQKPPRCVAFASDLIFATISLNFHREGRNSVAA
LSLDSHMFHPRADRWLLFVIVSPATSRATESRGFATGMENRKGVEVRCIEEE"
complement(20843..21366)
/gene="F6N23.20"
complement(join(20843..20944,21002..21056,21103..21155,
21292..21366))
/gene="F6N23.20"
/codon_start=1

```

```

/evidence=not experimental
/protein_id="AAC13621.1"
/db_xref="GI:3047110"
/translation="MVATAVASYFLLTGDYGPENVLDPYLSSMILLVEVLVIRMF
SYWMLHLFWLVIHWRKILAAQDSVQKQIFGPKGEPGKPEPSTAK"
complement(22410..24753)
/gene="F6N23.19"
complement(join(22410..22521,23220..23538,23620..23939,
24009..24137,24240..24352,24439..24548,24639..24753))
/gene="F6N23.19"
/codon_start=1
/evidence=not experimental
/protein_id="AAC13637.1"
/db_xref="GI:3047126"
/translation="MSPSYINVSILKPFPHLMQDILTNVVKYQGNWKRIAECLPG
SEENRNDVOQHRLKVLDPQLQKWKKEDELLSELVKDYMDNDRPWSKISKE
PGRKDCORRRHNLNPTIISKPTPEERELIQAQSGNGKWAETAKLLPGRTENN
IKHNGSVKRLKLEQFPFNIPLSGVYGSKSESSFEYFNQRYNMWESCITQKEAA
KSPORDFLDITLGLNWRKSISSISLURGEESVSSVSDVCARLNACLETQNNNDTV
CVKEVRMKRLRMARTFTPTSIISKTSPASGLKRLQKYDTPTPTDARSHMSBE
DHSVASPSKSYRFVKNTCSCGKPLERRLDFFLLWDEHGRNGIVNFSVRLPQKS
DLKSLVRRPFWLR"

```

Alignment Scores:

Pred. No.:	156e-85	Length:	91849
Score:	1116.50	Matches:	231
Percent Similarity:	68.38%	Conservative:	35
Best Local Similarity:	59.38%	Mismatches:	26
Query Match:	75.13%	Indels:	99
DB:	8	Gaps:	3

US-10-723-061-18 (1-292) x F6N23 (1-91849)

Qy 1 AlalysleleleaspGlyLysLeuValAlaLysGlnleAargluGluLeuAlaValGlu 20
 Db 48245 GCATTTGTAATAGATGAAAGGCTGTGCAAAAAGATTAGATGAAATCAATGAA 48186

Qy 21 IleAlalysMetLysaspAlalleGlyValValProGlyLeuAlaValleuValGly 40
 Db 48185 GTTTCAGATGAAGGAATCAATGTTGATCTCTGTTAGCAGTAATCTTGTGT 48126

Qy 41 SerArglyaspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
 Db 48125 GACAGAAAAGATTCTGCAACTTATGTGAGGACACAGAAAGCTTGTGACTCCGTGGA 48066

Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluaspGluValleuLysHis 80
 Db 48065 ATCAATCGTTCGAAGTTCGTCTAGCTGAAGATTCTATCAGAAAGAGGTGTGAAATCT 48006

Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
 Db 48005 GTATCAGGATTCAATGATGATCTCTGTCCTAGGAATCTTGTTCAGTTGCTCTGCCA 47946

Qy 101 His----- 101

Db 47945 TC-GGTAACAAAGTCTAATACTTTCTTCGGAACCTAAATAACACAGTGAATTACAATAG 47887

Qy 102 -----HisMetAsnAspGluAsnIleLeuAsnAlaValSerIle 114

Db 47886 TGTTTTGGGATGTTGTTGCGAGCATATGGATGAACAGACATACTGAAATCGGTTAGTATA 47827

Qy 115 GluLysaspValaspGlyPheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArg 134

Db 47826 GAGAAAGATGTTGACGGATTTCATCGCTAAATATTTGGACGCTTGCCATGCGTGAAGA 47767

Qy 135 AspProPhePheValProCysThrProLysGlyCysMetGluLeuLeuHisArgTyrGly 154
 Db 47766 GAACCCCTTATTCGTTCTTGTACTCCAAAGAGTTCGATGAGTTGTTGTCATAGATACAC 47707

Qy 155 ValGluIleLysGlyLysArgAlaValIleGlyArgSerAsnIleValGlyMetPro 174
 Db 47706 ATTGAATCAAGGAAAGAGCGGTTGTTTCGGAAGAGTAAACATTGTCGTTGCTGCA 47647

QY	175	AlaAlaLeuLeuLeuGln	-----	180
Db	47646	GCTGCTCTTTTACTGCAGGTAATGATGATGATACCTTCTTTAAAGATGCCGTTTTTCTCTGG	47587	
QY	180	-----	-----	180
Db	47586	TTGTGCACATTCGAAAGTTAGTACATGTTGGTCTTTTATTGTTGTGTTTGTATTGTTAG	47527	
QY	181	LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr	200	
Db	47526	AGGAGGATGCAACCGTTAGCATTTATCCATTCAGAACCAAGAACCTGGAAGAAATCACA	47467	
QY	201	ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTrp	220	
Db	47466	AGAGAAGCTGATATATAATCTCAGCTGTTGGACAGCCAAACATGGTTCAGAGGAGCTGG	47407	
QY	221	IleLysProGlyAlaAlaIleAspValGlyIleAsnProValAsp	-----	236
Db	47406	ATAAACCGGCGCAGTCCCTCATCGATGTTGGGATTAATCTCTGTGA-GGTGAGTGAAT	47348	
QY	236	-----	-----	236
Db	47347	TTTCTTTTGTACATATGCTCATGCGGTAATTCACTGATATATAAATGAATTTCAAATCAACCA	47288	
QY	237	-----	-----AspProGluSerProAtsGly	243
Db	47287	GTATGCAACATTCATACATGATACATATAATGAATTCAGGATCCAAAGTGTGCCCGTGA	47228	
QY	244	TyrArgLeuValGlyAspValIleCysTyrGluGluAlaSerLysIleAlaGlyLeuIleThr	263	
Db	47227	TATCGATTGTTGGAGACATTTGTCACGAGGAGCTAGCAAGTTGATCAGCCATCACA	47168	
QY	264	ProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsnThrLeuGlu	283	
Db	47167	CCTGTCTCTCGCGGTGAGGACCAATGACCATGACCATGCTTCTATCCAACTTTAACA	47108	
QY	284	SerAlaLysArgIleHisLysPheLys 292		
Db	47107	TCAGCTAAGAGATTCACAACTTCAG 47081		
RESULT B				
ATCHRIV2				
LOCUS	197975 bp	DNA	linear	PLN 16-MAR-2000
DEFINITION	Arabidopsis thaliana DNA chromosome 4, contig fragment No. 2.			
ACCESSION	AL161472			
VERSION	AL161472.2	GI:7267383		
KEYWORDS	Arabidopsis thaliana (thale cress)			
SOURCE	Arabidopsis thaliana			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE	1	(bases 1 to 197975)		
AUTHORS	Lamar, B., Stoneking, T., Stumpf, J., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.			
JOURNAL	Unpublished			
REFERENCE	2	(bases 1 to 197975)		
AUTHORS	EU Arabidopsis sequencing, project.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk			
COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV1 at the 5' end and an overlap with ATCHRIV3 at the 3' end.			
FEATURES	Location/Qualifiers			
source	1..197975			
	/organism="Arabidopsis thaliana"			

```
/mol_type="genomic DNA"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
10352..17547
/gene="AT4g00450"
join(10352..10384,10432..10545,10605..10709,10849..11018,11079..11423,11475..13674,13735..13823,13921..16530,16677..16862,16976..17042,17092..17547)
/gene="AT4g00450"
/note="Contains Aminoacyl-transfer RNA synthetases class-II signatures AAL343-1352;Prokaryotic membrane lipoprotein lipid attachment site AA349-359 contains EST gb:AI999580.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB80854.1"
/db_xref="GI:7267384"
/db_xref="SPTREMBL:O23057"
/translacion="MLSPFLVISIGAVKCLRAINESRALKRKVKIKRLSDSNPFOFIFCAEPRAGQVGYVPLSGSLICKPGPEQSCGEETKKRWTESLSQOKHSLADINIPGYARKTLFEVLRNNVPLLRATWPIKVTYLNQVHCWAINWCDLVLLFLQEHLTRHLLKLVYSNGQMLLNICNTSNMNFCHGTAHLLSKLEIGHRCFIQDQCKRIVQQAQATARKELYTINGICWCFYSGTLMKGFFELISLLIGFSSSYFVNFHFAHLLKEIFELQLLLPIVGVLESIVLSQTVQSLAVAVRFIQEAPFGSDLVNRRAYTLEISALIMVRLVLAAPTFVASDFPPLPPVAACGPNDSYTSKAYENLEKLSRAISDAIQFQGRYVLRPEFLSFDYTIETIORSADDLAKIASAGYPOHNAKAVQALKALSDGIDRAAYSYLFDLNGAVDEAMITVSPCLRSRLWIGAISTFVCSVFFLEWATCDPDRFRAGVPKIDFKSGKDCSQVLYLQLLKXKLGEBTARKGKCNRFVGVKPSGSMDAFEISGPHDIIVWIDQEHVHGKAKRLQLLLELIRSGIFNPATYVRQLIVSMIDVIQPAVDPERNRHRIILKQIPGCFVHETLEEAQLFGDKLJSEAVRTYSNERRLLRELLYKXGKYNNLIVLSDQSKKISTLSVIFPACNAKSNKSPKRTKSSYDIBELKERISALLQPGMSGVTPEVDFONSVKRSQSVYVKMDQPEATPGCEDCRARPKMNDKSSQVQGNSTADEEDNMWTKGSKTVESLKVDPQIEITKVQPRGQKMARKTQSLAQONARLEGSGASTSHVCNDKYSCHPHGPGVEGOKVVDVFTSTPVMVSGNLSKQCFVQKRSIAVWLITTAVALQVLEEPQKSSVRVQGNRGAPVEKNTIRWKLGADELYSILFLDIDLVLVAGGNLTVPRNVNNEICEIGALLVSSLRRYENIILSADLYPEAMTLMNRAASLMSSNGKISGAALVTVYILKRYGLSPSVWHNNFKATSEKKLLSELDHTRSGNGEYGNPLGVPAGVDNDPDLRKKISIGGARPSRVGLSNRDLQRHEEATHYLKLTGTWKASLAEKNDQYQVQVIVGLMDCIRQTGGAQEGPPLVSSAVSAIINSVGLSVARITDFSLGNIYQNHPSGVDNSNIARYILRHITCLLKEALGERVPEIATATESSTALTGVFAPVKVSGHQHQLSPESVDSNANNSTIDMSNGTGAWALSRATKITAASVAVIGSTHGVITLIRVIGLLRLKDYLDVQVFRTRTSSNGSARGKAGVESPIEYVYVHVRLLVNGCTVSEGLVLELVGESSVVAISRQKMLPLKLVFPAYSIIFAVLMRFPVSNSSNSVHEDTHLYQSLTMADHVKHLPFRDVCFRDTQGLYELIIVADSTDAEFASFVESHGDLHLKSAFAPLARLFLNSLDCKVPVSGYSHGVSEAKNRHOGNGTKLVDLVSLVCLQPAKFWQVVELRLLEQAALAEKLEHNDPLTQAIRSSCPTSEKPDASENKNFIOILLTRLLVRPDVPLFVSEVHLFGKRSVEDSMLKQAEWFLAGQDVLFGKRTIRQKLIIVGESKGLTKPQFWKPGWKNSSSDHILITANKAGKXKPEITISIEGEVIEEGSKVLLPRVLDENSFSGYGITTERAFVQLVLPCTDQSDSESTFVNLVRFQSNIEQOLSVTNRTSTTNKQMTAGSGISSEKSGSTRKGLRGSPSLARRSSANTTDTSPPSPALRASMLRLQFLRLLPVICGEPSFKNTRHALASTIVRLGSRVVYEDYAVCSPRSELSKAETETIDPSSMADLSSELVDFRELTGLSTNFKSLKVCLLGCKNELSRMOLPDTIRWRIQAAKPTILLPSLRCSLSCPHSVPTPALTLVQPSGSTAAAGTQNRNSPAISKGTAAQGLKPTMLAPHQOQCADNTDVVDPWTLLEDTSSGLSSSNASNSDVAHLRATCWLGKAVRVRRTLTYVGSVDDDS"
10352..10384
/gene="AT4g00450"
/number=1
10385..10431
/gene="AT4g00450"
/number=1
10432..10545
/gene="AT4g00450"
/number=2
10546..10604
/gene="AT4g00450"
/number=2
10605..10709
/gene="AT4g00450"
/number=3
10710..10848
/gene="AT4g00450"
```

gene

CDS

exon

intron

exon

intron

exon

intron


```
RESULT 11
AY085538 1182 bp mRNA linear PLN 14-APR-2003
LOCUS
DEFINITION Arabidopsis thaliana clone 156450 mRNA, complete sequence.
ACCESSION AY085538
VERSION AY085538.1 GI:21404248
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1182)
Haas, B.J., Volkovskiy, N., Town, C.D., Troughan, M., Alexandrov, N.,
Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. 3 (6), RESEARCH0029 (2002)
22088475
PUBMED 12093376
2 (bases 1 to 1182)
Brovet, V., Troughan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1182)
Brovet, V., Troughan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and
Feldmann, K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the WS or Laer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.
FEATURES
Location/Qualifiers
1..1182
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="156450"
90..989
/codon_start=1
/product="5,10-methylenetetrahydrofolate
dehydrogenase:5,10-methylenetetrahydrofolate
cyclohydrolase, putative"
/protein_id="AA062762.1"
/db_xref="GI:21553669"
/translation="VASSSDHTAKIIDGKAIAHTIRSEAEVRLGSEKHKGVPLAV
VIVGRKDSOTVNTKRVKAEVGIKSFVGLPEVSEADLIISKVHLSNPDVHGIL
VOLPFPKHNENIGALISIDKVDGPHPLNIGLAKMGREPLPLCTPKGCELLAR
SGVKIKGRVAVVGRSNIVLPSVLLLRADATVTVVSHTKQPEALIREADIIVAAC
GQAHMKNWIKPGAIVDVGTVSPKSGYRLVDGVDFAEASKVAGFITPVPVG
VGEPTVAMLLNTVDGAKRVGE"
```

ORIGIN

Alignment Scores:
Pred. No.: 1,01e-80 Length: 1182
Score: 1035.00 Matches: 194
Percent Similarity: 82.99% Conservative: 45
Best Local Similarity: 67.36% Mismatches: 49

```
Query Match: 69.65% Indels: 0
DB: 8 Gaps: 0
US-10-723-061-18 (1-292) x AY085538 (1-1182)
QY 1 AlaLysIleAlaAspGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
DB 114 GCGAGATAATCGACGGCAAGCGGATTGCTCATCATCAGATCAGATCCCGAGGAA 173
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLysLeuAlaValIleLeuValGly 40
DB 174 GTTCGCGGTCTATCTCAGAAACACCGCAAGGTCGCCAGGACTAGCTGTAGTTATTGTGGG 233
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
DB 234 AGCCGAAGGATTCACAGACCTATGTGAATACGAAGAGGAAAGCGTCGCTGAGGTGG 293
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
DB 294 ATTAAGTCATTTGACGTGGCCCTCCAGAGGAAGTTTCTGAAGCTGATCTTATTAGCAA 353
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
DB 354 GTTCATGAACATAATCAATCGGATGTCCATGGCATATTAGTTCACATCCCATTCGCG 413
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
DB 414 AAACATATTATAGGAGAAATATATTGGTGCAATCAGCATTCATAAAGATGTCGATGC 473
QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
DB 474 TTCATTCCTTTGAATATTGGTAAGCTAGCCATGAAGGAGGAGAACCCCTCTTCCTCT 533
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyHis 160
DB 534 TGCAACCCCAAGGAGTGTGGAACTCTAGCTAGAACGGCGGTAAAGATAAAGGGCAA 593
QY 161 ArgAlaValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
DB 594 CGAGCAGTGTGTGTAGCTCGAGTAACATTTGTGGTTTCCCGCTTCACATCTTTTGTCT 653
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
DB 654 AAGCTGTAGCTACTGTACACTGTACATCTCACACCAAGGATCTCGAGGCTATCATA 713
QY 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTyr 220
DB 714 CGGGAAGCTGACATTTGTTTGTCTGTCGCGACACAGCCCATGATTAAAGGCAACTGG 773
QY 221 IleLysProGlyAlaAlaIleIleAspValGlyIleAsnProValAspAspProGluSer 240
DB 774 ATAAAGCCAGGCGCTGCAGTAATTGATTTGGNACTAATGCAGTCAGCGACCCGAGCAAG 833
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
DB 834 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCGCAGAAGCTTCAAAAGTTGAGGT 893
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
DB 894 TTCATAACTCCGTCCTCGTGTGTAGGCCCAATGACAGTGGCAATGCTTCTCAGGAAC 953
QY 281 ThrLeuGluSerAlaLysArgIle 288
DB 954 ACCGTAGACCGTGCCCAAGCGGTGTC 977
RESULT 12
AF030516 1219 bp mRNA linear PLN 05-JAN-1999
LOCUS
DEFINITION Pisum sativum 5,10-methylenetetrahydrofolate
dehydrogenase-5,10-methylenetetrahydrofolate cyclohydrolase mRNA,
complete cds.
ACCESSION AF030516
VERSION AF030516.1 GI:4103986
```

KEYWORDS SOURCE ORGANISM

Pisum sativum (pea)
Pisum sativum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
Pisum.
1. (bases 1 to 1219)
Chen, L., Nargang, N.E. and Cossins, E.A.
a cDNA sequence encoding pea leaf cytosolic bifunctional
5,10-methylenetetrahydrofolate
dehydrogenase-5,10-methylenetetrahydrofolate cyclohydrofolase
Unpublished
2. (bases 1 to 1219)
Chen, L., Nargang, N.E. and Cossins, E.A.
Direct Submission
Submitted (20-OCT-1997) Chemistry and Biochemistry, University of
Texas at Austin, Austin, TX 78705, USA
JOURNAL
Location/Qualifiers

REFERENCE AUTHORS TITLE

1. 1219
Chen, L., Nargang, N.E. and Cossins, E.A.
Isolation and sequencing of a cDNA encoding pea cytosolic
bifunctional 5,10-methylenetetrahydrofolate
dehydrogenase-5,10-methylenetetrahydrofolate cyclohydrofolase.
Comparisons with the yeast, mammalian, and bacterial proteins
Unpublished
2. (bases 1 to 1219)
Chen, L.
Direct Submission
Submitted (28-SEP-1998) Chen L., Biochemistry and Chemistry,
University of Texas at Austin, Speedway and 24th St., Austin, Texas
78731, USA

FEATURES source

1. 1219
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
68..952
/note="bifunctional protein; 31.3 kDa protein"
/codon_start=1
/product="5,10-methylenetetrahydrofolate
dehydrogenase-5,10-methylenetetrahydrofolate
cyclohydrofolase"
/protein_id="AAD01907.1"
/db_xref="GI:4103987"
/translations="MATVDIGKVAQTIIRSEADEVILRSQYKGVPLAVVIGNRK
DSQSYVMKRAKAEIGIKSFIDLPEDASEAEIKNVHELNANPDVGLVLPK
HUNERKVLKLSKVDGPHPLNGLKAMGRDPLFPCPKAGLELSRSGVSIK
KXAVVGRNINVLPSLLKADATVIVHSHTSQPFTIIRADIVITAAAGQAKWIK
GSWIKFQAVINDVETSDPTIRKSGYRLVGDVDFEASKVAGWITPVGVGPMVT
MLLKNTLEGAKRTIEQN"

CDS

ORIGIN

Alignment Scores:
Pred. No.: 1 28e-80 Length: 1219
Score: 1034.00 Matches: 193
Percent Similarity: 82.93% Mismatches: 45
Best Local Similarity: 67.25% Indels: 0
Query Match: 69.58% Gaps: 0
DB:
US-10-723-061-18 (1-292) x AF030516 (1-1219)

QY 1 AlaLysIleAlaSpGlyLysLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
DB 71 GCCACCGTAATGACGGCAAGCGTTGACAACTATCCGATCGAATCCGCCAGCAG 130
QY 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
DB 131 GTTCGCTTCTCCAAAGTACGCGAGGTTCTCGACTTGGAGTGTGATAGTAGA 190
QY 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysLysAlaCysGluAlaValGly 60
DB 191 AACCGAAGGACTCTCAAAGTACGTTGGAATCAAGAGAAAGCGTGTCTGAATGGGA 250
QY 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
DB 251 ATCAATCTTCTGATATGACCTTCGGAGGATGCTTCGAAGCTGAATCAATAAAAT 310
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
DB 311 GTTCAGGAATTGAACGCTAACCCGATGTACATGTATATTGTTCACTTCATTCGCT 370
QY 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
DB 371 AAGCAGTAATGAAGAGAAAGTTTGTACTGAATCATGCAATTCAAAGGATGTAGATGC 430

QY 121 PheHisProLeuAsnIleGlyArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
DB 431 TTCCATCCCTTGAACATTTGGCAAGCTTGCATGAAAGGAAGAGACCTCTGTTTCTCCA 490
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
DB 491 TGTACTCCCAAGGCGATGCTTTGAACATTTATCAGCAAGTGGTGTAAAGTATATAAGGGGAAA 550
QY 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
DB 551 AAGGCTGTGTGGTGTAGAGCAACATAGTTGGATTACCAAGCTTCAATGCTCTCTG 610
QY 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
DB 611 AAAGCAGATGTACAGTTACCATTTGTTCAATTCACACACCAAGTCAACGAGAACTATCAT 670
QY 201 ArgGlnAlaAspIleValIleAlaAlaValGlyValAlaAsnLeuValArgGlySerTyr 220
DB 671 CGTGAAGCAGATATTTGTTATTCAGCAGCAGCAGCAAGATGATCAAGGAGGAGCTGG 730
QY 221 IleLysProGlyAlaAlaIleAlaAspValGlyIleAsnProValAspProGluSer 240
DB 731 ATAAACCCAGGAGCTGCAGTATAGATGTTGGCACAATTTCTGTGGATGACCCCACTAGG 790
QY 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
DB 791 AAATCAGGTTATAGACTTTGTTGGAGATGTAGATTTTGAGGAAGCATCTAAAGTTGCTGGT 850
QY 261 LeuIleThrProValProGlyGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
DB 851 TGGATTACTCTGTTCTGTTGGTGGTCCCAATGACACTCACAATTTGTTGCTGAAGAAT 910
QY 281 ThrLeuGluSerAlaLysArg 287
DB 911 ACTTTGAGGCGCGCTAAACGC 931

RESULT 13

PSA011589 1219 bp mRNA linear PLN 29-SEP-1999
LOCUS Pisum sativum mRNA for bifunctional 5,10-methylene-
tetrahydrofolate dehydrogenase and 5,10-methylenetetrahydrofolate
cyclohydrofolase protein.
DEFINITION
AJ011589
ACCESSION
VERSION AJ011589.1 GI:6002382
KEYWORDS 5,10-methylenetetrahydrofolate cyclohydrofolase;
5,10-methylene-tetrahydrofolate dehydrogenase; bifunctional enzyme.
SOURCE Pisum sativum (pea)
ORGANISM Pisum sativum

REFERENCE

1. Chen, L., Nargang, F.E. and Cossins, E.A.
Isolation and sequencing of a cDNA encoding pea cytosolic
bifunctional 5,10-methylenetetrahydrofolate
dehydrogenase-5,10-methylenetetrahydrofolate cyclohydrofolase.
Comparisons with the yeast, mammalian, and bacterial proteins
Unpublished
2. (bases 1 to 1219)
Chen, L.
Direct Submission
Submitted (28-SEP-1998) Chen L., Biochemistry and Chemistry,
University of Texas at Austin, Speedway and 24th St., Austin, Texas
78731, USA

FEATURES

source
1. 1219
/organism="Pisum sativum"
/mol_type="mRNA"
/db_xref="taxon:3888"
68..952
/EC_number="1.5.1.5"
/note="bifunctional enzyme"
CDS

```

/codon_start=1
/evidence=experimental
/product=5,10-methylenetetrahydrofolate dehydrogenase:
5,10-methylenetetrahydrofolate cyclohydrolase"
/protein_id="CAB56756.1"
/db_xref="GI:6002383"
/db_xref="GOA:Q92TV0"
/db_xref="SPTREMBL:Q92TV0"
/translation="MATVIDGKVAQTIRSEIADVEVRLLSOKYGVKPLAVLVGNRK
DSQSVGMKRCACALGKSFIDLPEDASEAELIKNVHELNANPDVHGILVOLPLPK
HYNKEVLTEISIKDVGDPHPLNIGLANKGRDPLPCTPACLELLSRGSVSIKG
KRAVGRNSRIVGLPASLLKADATVIVHSITSQETIIRADIIVAAAGQAKIK
GSWIRPGAAVIDGTVNSDPTKRSYRLVDVDFEASKAVAGMITPVPGGVPMVT
MLLKXNLEGAKRTEQN"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 1,28e-80 Length: 1219
Score: 1034.00 Matches: 193
Percent Similarity: 82.93% Conservatives: 45
Best local Similarity: 87.25% Mismatches: 49
Query Match: 69.58% Indels: 0
DB: 8 Gaps: 0

```

US-10-723-061-18 (1-292) x PSA011589 (1-1219)

```

Qy 1 AlaLysIleLeuAspGlyLeuValAlaLysGlnIleArgGluGluIleAlaValGlu 20
Db 71 GCCACCGTAATCGACGCAAGCGGTTCACAAAATATCCGATCTGAAATCGCCGACGAG 130
Qy 21 IleAlaLysMetLysAspAlaIleGlyValValProGlyLeuAlaValIleLeuValGly 40
Db 131 GTTCCTCTCTCACAAGATACGCAAGGTTCCTGACATTCGAGTTCGATAGTAGGAG 190
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValGly 60
Db 191 AACCGAAAGGACTCTCAAGCTACGTTGCAATGAAGAAAGCGTGTGCTGAATCGGA 250
Qy 61 IleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 251 ATCAATCCTTCGATATCGACCTTCGAGAGATGCTTCGAGTGAATCAATAAATAAT 310
Qy 81 IleAlaThrPheAsnSerAspProSerValHisGlyIleLeuValGlnLeuProLeuPro 100
Db 311 GTTCACGAAATGAAGCGTAAACCCGATGATGATGATGATGATGATGATGATGATG 370
Qy 101 HisHisMetAsnAspGluAsnIleLeuAsnAlaValSerIleGluLysAspValAspGly 120
Db 371 AAGCAGTAAATGAAGAAAGATTTTGACTGAAATCAGCAATTCAGAGGATGTAGATGC 430
Qy 121 PheHisProLeuAsnIleGlyValArgLeuAlaMetGlnGlyArgAspProPhePheValPro 140
Db 431 TTCCATCCCTTGACATTTGGCAGCTTGCATGATGAAGAAAGAGACCTCTGTTCTTCCA 490
Qy 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrGlyValGluIleLysGlyLys 160
Db 491 TGTACTCCAGGCGATGCTTCAACTATTATCACCAGTGTGTGAAGTATAAGGGGAAA 550
Qy 161 ArgAlaValValIleGlyArgSerAsnIleValGlyMetProAlaAlaLeuLeuGln 180
Db 551 AAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 610
Qy 181 LysAlaAsnAlaThrValSerIleValHisSerAsnThrLysLysProGluGluIleThr 200
Db 611 AAAGCAGATGCTACAGTTACCATTTCTTATTCACACACACAGTCAACAGAACTATCAT 670
Qy 201 ArgGlnAlaAspIleValIleAlaValGlyValAlaAsnLeuValArgGlySerTrp 220
Db 671 COTGAAGCAGATATTGTTATTTCACGACGACGACGACGACGACGACGACGACGACG 730
Qy 221 IleLysProGlyAlaAlaIleLeuAspValGlyIleAsnProValAspProGluSer 240
Db 731 ATAAACACGAGGAGCTGCTAGTGTAGTGTGGCACAATTTCTGGATGACCACTAGG 790

```

```

Qy 241 ProArgGlyTyrArgLeuValGlyAspValCysTyrGluGluAlaSerLysIleAlaGly 260
Db 791 AAATCAGGTTATAGACTTTTGGAGATGTAGATTTGAGAAAGCATTAAGATTCGTGT 850
Qy 261 LeuIleThrProValProGlyValGlyValGlyProMetThrIleAlaMetLeuLeuSerAsn 280
Db 851 TGGATTACTCTCTGTTCTGCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 910
Qy 281 ThrLeuGluSerAlaLysArg 287
Db 911 ACTTTGAGGGGCGCTAAACGC 931

```

RESULT 14

```

BT004612 LOCUS 900 bp mRNA linear PLN 19-FEB-2003
Arabidopsis thaliana At3g12290 gene, complete cds.

```

```

BT004612
BT004612.1 GI:28416654
FLI CDNA.
Arabidopsis thaliana (thale cress)

```

```

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

```

REFERENCE

```

AUTHORS
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
Ecker,J.R.

```

TITLE

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 900)

REFERENCE

```

AUTHORS
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Chan,M.M., Chang,C.H., Dale,J.M., Hayashizaki,Y., Hsuan,V.W.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K.,
Yu,G., Yuan,S., Shinozaki,K., Davis,R.W., Theologis,A. and
Ecker,J.R.

```

Direct Submission

```

Submitted (19-FEB-2003) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

```

COMMENT

```

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

```

The Salk, Stanford, PGECC (SSP) Consortium members constructed and
sequenced the pUNI (ORF) clones using the RAFL cDNAs: Cheuk,R.,
Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M.,
Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B.,
Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J.,
Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and
Ecker,J.R.

```

```

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.

```

FEATURES

source

```

1..900
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"

```

```

/chrosome="3"
/clone="U22770"
/ecotype="Columbia"
/notes="This clone is in PUN1 51"
1. 900
CDS
/notes="5,10-methylenetetrahydrofolate
dehydrogenase; 5,10-methylenetetrahydrofolate
cyclohydrolase, putative."
/codon_start=1
/product="At3g12290"
/protein_id="AA042858.1"
/db_xref="GI:28416655"
/translation="MASSSDHTAKIDGKAIHTIRSEIAEEVRLSKHGKVPGLAV
VIVGSRKDSQTYVNTKRKCAEVIKSFVGLPEEVSADLI SKVHEINSPDVHGIL
VOLPLKINEEHLICAGIDKVDGPHLNI GKLAMKREPLFPCTPKGCLLELLAR
SGVKIKGQRAVVGRSNI VGLPVSLLLKADATVTVSHSTKOPRAIIRADIVTAAC
GQAHMIKGNWIKRPAVIDVGNVSDPSKSGYRLVGDVDFPAEASKVAGTTPVPGG
VGPMIVAMLRNTVDGAKRVGE"

ORIGIN
Alignment Scores:
Pred. No.: 1.98e-80 Length: 900
Score: 1030.00 Matches: 193
Percent Similarity: 82.9% Conservatives: 46
Best Local Similarity: 67.01% Mismatches: 49
Query Match: 69.31% Indels: 0
DB: 8 Gaps: 0
US-10-723-061-18 (1-292) x BT004612 (1-900)

QY 1 AlaAlaValleAspGlyLeuValAlaLysGlnLeuArgGluLeuAlaValGlu 20
Db 25 GCAGAGATATTCACGCGCAAGCGATGTCTATACATCAGATCAGATCCCGCAGGAA 84
QY 21 IleAlaLysMetLysAlaValleGlyValProGlyLeuAlaValleLeuValGly 40
Db 85 GTTCGGGTCTATCTAGAACACGCGCAAGTCCAGACTAGTCTAGTTATTGTGG 144
QY 41 SerArgLysAspSerGlnThrTyrrValArgAsnLysLysLysAlaCysGluAlaValGly 60
Db 145 AGCCGAAGGATTCACAGACCTATGTGAATACGAGAGAAAGCGTCGCTGAGGTGG 204
QY 61 IleLysSerTyrrGluValAsnLeuProGluAspSerSerGluAspGluValLeuLysHis 80
Db 205 ATTAAGTCATTGACGTGGCTACAGAGGAAGTTCTGAAGTCATCTATTAGCAA 264
QY 81 IleAlaThrPheAsnSerAspProSerValHisGlyLeuValGlnLeuProLeuPro 100
Db 265 GTTCATCACTAAATCAATCCGATGTCATGTCATGATATTAGTTCAACTCCATTGCG 324
QY 101 HisHisMetAsnAspGluAsnLeuLeuAsnAlaValSerLysGluLysValAspGly 120
Db 325 AACATATTATGAGAGACATATTGCGTGCATCAGCATTTGATTAAGATGTCATGCC 384
QY 121 PheHisProLeuAsnLysGlyArgLeuAlaMetGlnGlyArgAspProPheValPro 140
Db 385 TTCATCCTTTGAATATTGTAAGCTAGCAGTCAAGAGGAGGAGGAGGAGGAGG 444
QY 141 CysThrProLysGlyCysMetGluLeuLeuHisArgTyrrGlyValGluLysGlyLys 160
Db 445 TGACCCCAAGAGGATGTTTGAACCTCTAGCTAGAGCGGCTTAAGATAAGGAGG 504
QY 161 ArgAlaValleValleGlyArgSerAsnLeuValGlyMetProAlaAlaLeuLeuGln 180
Db 505 CGAGCAGTGTGTAGTCCGAGTAACTATTGTTGTTTCCCGTTTCACTCTTTTGTC 564
QY 181 LysAlaAsnAlaThrValSerLysLeuHisSerAsnThrLysLysProGluLeuLeuThr 200
Db 565 AAGCGTGTACTGTCTCAACTGTACATCTTCACCAAGATCCCTGAGGCTATCATA 624
QY 201 ArgGlnAlaAspLeuValleAlaValGlyValAlaAsnLeuValArgLysSerTrp 220
Db 625 CGGGAAGCTGACATTGTTTATGTCATCGGACCAAGCCACATGATTAAAGGCAACTGG 684

```

```

QY 221 IleLysProGlyAlaAlaValleLeuAspValGlyLeuAsnProValAspAspProGluSer 240
Db 685 ATAAAGCCAGGGCTGCAGTAATGATGTTGGAACTAATGCACTCAGCCAGCAGCAAG 744
QY 241 ProArgGlyTyrrArgLeuValGlyAspValCysTyrrGluGluAlaSerLysLeuAlaGly 260
Db 745 AAATCAGGATACCGGTTGGTTGGAGATGTTGATTTCCGCAAGCTTCAAAAGTTGCAGGT 804
QY 261 LeuLeuThrProValProGlyGlyValGlyProMetThrLeuAlaMetLeuLeuSerAsn 280
Db 805 TTCATAACTCCGTCCTCGTGGTGTAGGCCCAATGACATGCGCAATGCTTCTCAGGAAC 864
QY 281 ThrLeuGluSerAlaLysArgIle 288
Db 865 ACCGTAGACGGTCCCAAGCGTGC 888

RESULT 15
AKI03505
LOCUS
DEFINITION
Oryza sativa (japonica cultivar-group) cDNA clone:J033131D02, full
insert sequence.
ACCESSION
AKI03505
VERSION
AKI03505.1 GI:32988714
KEYWORDS
FLI_CDNA; CAP trapper.
SOURCE
Oryza sativa (japonica cultivar-group)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team;
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Otsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group; Otono, Y., Murakami, K.,
Ida, I., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,
Kusumi, T., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Ikumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M. and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)
JOURNAL
MEDLINE
22752273
PUBMED
12869764
2 (bases 1 to 1118)
AUTHORS
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Ida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kawagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sasaki, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,

```

Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression: 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-338-7007, Fax:81-29-338-7007)
This clone is one of the 28K full-length cDNA clones from japonica rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yanagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Otomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Niikura,J., Oka,M., Ryu,N., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Iihaka,T., Hori,F., Iida,Y., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Maezuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Satoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,T., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1. 1118
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033131D02"

ORIGIN

Alignment Scores:
Pred. No.: 9,45e-79 Length: 1118
Score: 1012.00 Matches: 191
Percent Similarity: 80.84% Conservative: 41
Best Local Similarity: 66.55% Mismatches: 55
Query Match: 68.10% Indels: 0
DB: 8 Gaps: 0

US-10-723-061-18 (1-292) x AK103505 (1-1118)

Qy 1 AlalysllelleAspGlyLysLeuValAlalysGlnlleArgGluGluIleAlaValcglu 20
Db 117 GCGCATGATCATCGACGGAGGCGGTGCGCGCCACATCCGCCGAGGTGCGCGCCGAC 176
Qy 21 lleAlalysMetLysAspAlaalleGlyValValProGlyLeuAlaValleLeuValcgl 40
Db 177 GTCCGCCGCCCTCTCTCTCCGCCCAACACTCGTGTCCCGCGGTGCGCGGTGATCGTGGG 236
Qy 41 SerArgLysAspSerGlnThrTyrValArgAsnLysLysAlaCysGluAlaValcgl 60
Db 237 AGCAGGAGGACTCGACAGCTAGTCAGATGAAGCGCAAGCGTGGCCGAGGTGGC 296
Qy 61 lleLysSerTyrGluValAsnLeuProGluAspSerSerGluAspGluValleuLysHis 80
Db 297 ATCCGCTCCGTCGACGTGACCTCGCCGAGACATCTCCGAGGGGGGGCGTCTGTCGCCGAG 356
Qy 81 lleAlaThrPheAsnSerAspProSerValHisGlylleLeuValGlnLeuPro 100
Db 818

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2004, 06:09:35 ; Search time 84 Seconds
(without alignments)
1096.801 Million cell updates/sec

Title: US-10-723-061-18

Perfect score: 1486

Sequence: 1 AKIDGKLAKQIREIAVE.....TIAMLLSNTLESAKRIHKF 292

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 31518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_ARCHAEA.*

2: SP_BACTERIA.*

3: SP_FUNGI.*

4: SP_HUMAN.*

5: SP_INVERTEBRATE.*

6: SP_MAMMAL.*

7: SP_MHC.*

8: SP_ORGANELLE.*

9: SP_PHAGE.*

10: SP_PLANT.*

11: SP_RODENT.*

12: SP_VIRUS.*

13: SP_VERTEBRATE.*

14: SP_UNCLASSIFIED.*

15: SP_RVIRUS.*

16: SP_BACTERIAP.*

17: SP_ARCHAEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1238	83.3	360	10 065271	065271 arabidopsis
2	1035	69.7	299	10 081EA3	081ea3 arabidopsis
3	1034	69.6	294	10 09ZTV0	09ztlv0 pisum sativ
4	1030	69.3	299	10 09LHH7	09lhn7 arabidopsis
5	1017	68.4	350	10 09ZVH8	09zvh8 arabidopsis
6	1009	67.9	310	10 065269	065269 arabidopsis
7	992.5	66.8	291	10 09FRB2	09frb2 oryza sativ
8	792.5	53.3	160	10 08W527	08w527 zea mays
9	765.5	51.5	937	3 09HFV0	09hfv0 yarrowia li
10	763.5	51.4	284	16 09ZB24	09zb24 listeria in
11	755.5	50.8	284	16 08Y7C5	08y7c5 listeria mo
12	749.5	50.4	937	3 08WZJ7	08wzj7 schizosacch
13	743.5	50.0	284	16 099YX2	099yx2 streptococ
14	743.5	50.0	284	16 08K6S8	08k6s8 streptococ
15	741	49.9	299	16 08FV26	08fvp6 brucella su
16	741	49.9	319	16 08YCL8	08ycl8 brucella me

ALIGNMENTS

RESULT 1

O65271 PRELIMINARY; PRT; 360 AA.
 ID O65271
 AC O65271
 DT 01-AUG-1998 (TRENDELrel. 07, Created)
 DT 01-AUG-1998 (TRENDELrel. 07, Last sequence update)
 DT 01-OCT-2003 (TRENDELrel. 25, Last annotation update)
 DE F6N23.26 protein (Putative tetrahydrofolate synthase) (Hypothetical protein).
 DE F6N23.26 OR AT4G00620.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA WASHU;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Geisel C.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Waterston R.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RL [4]
 RC SEQUENCE FROM N.A.
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL [5]
 RC SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

17 740.5 49.8 934 13 Q7ZXF4
 18 738 49.7 725 11 Q8BXX7
 19 738 49.7 935 11 Q8R013
 20 730 49.1 344 16 Q9A8Y0
 21 724 48.7 285 16 Q97RI9
 22 721 48.5 284 16 Q8EGM3
 23 721 48.5 284 16 Q8E168
 24 716 48.2 279 16 Q9K966
 25 715 48.1 299 16 Q989A7
 26 715 48.1 306 16 Q8DD3
 27 714 48.0 284 16 Q8DVC1
 28 709.5 47.7 285 16 Q8EQ43
 29 707.5 47.6 294 16 Q9CH85
 30 704.5 47.4 281 16 Q836W7
 31 704 47.4 284 16 Q9JWI9
 32 703 47.3 932 5 Q27772
 33 701.5 47.2 291 16 Q7VVM3
 34 699.5 47.1 283 16 Q7WJG1
 35 699.5 47.1 283 16 Q7WAC0
 36 699.5 47.1 283 16 Q7VZC8
 37 698.5 47.0 291 16 Q7W4Z7
 38 694 46.7 294 16 Q89WX9
 39 689 46.4 282 16 Q9PP68
 40 689 46.4 299 16 Q92RY3
 41 689 46.4 327 16 Q7UNN9
 42 685.5 46.1 289 16 Q8XZ10
 43 685 46.1 284 16 Q8EG21
 44 684.5 46.1 286 16 Q818R5
 45 680.5 45.8 289 16 Q88WM8

Q7zxf4 xenopus lae
 Q8bxx7 mus musculu
 Q8r013 mus musculu
 Q9a8y0 caulobacter
 Q97ri9 streptococ
 Q8egm3 streptococ
 Q8e168 streptococ
 Q9k966 bacillus ha
 Q989a7 rhizobium l
 Q989a7 rhizobium l
 Q8d9q3 streptococ
 Q8dvc1 streptococ
 Q8eq43 oceanobacil
 Q9ch85 lactococcus
 Q836w7 enterococcu
 Q9jwi9 neisseria m
 Q27772 s c-1-tetra
 Q7vwm3 bordetella
 Q7wjg1 bordetella
 Q7wac0 bordetella
 Q7vzc8 bordetella
 Q7w4z7 bordetella
 Q89wx9 bradyrhizob
 Q9pp68 campylobact
 Q92ry3 rhizobium m
 Q7unn9 rhodopirell
 Q8xz10 ralsconia s
 Q8eg21 shewanella
 Q818r5 bacillus ce
 Q88wm8 lactobacill

```

[6]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF058919; AAC13627.1; -.
DR EMBL: AL161472; CAB80871.1; -.
DR EMBL: AY059944; AAL24426.1; -.
DR EMBL: BT000066; AAN15385.1; -.
DR PIR: T01226; T01226.
DR KSSP: P11586; 1A41.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0003936; P:folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR000672; THF_DHG_CyH.
DR Pfam: PF00763; THF_DHG_CyH; 1.
DR PRINTS: PR00085; THFDHGRNASE.
DR ProDom: PD002300; THF_DHG_CyH; 1.
DR PROSITE: PS00767; THF_DHG_CyH_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 360 AA; 38741 MW; 15D254E1DD2F6B2 CRC64;

Query Match 83.3%; Score 1238; DB 10; Length 360;
Best Local Similarity 79.1%; Pred.No. 1.1e-95;
Matches 231; Conservative 36; Mismatches 25; Indels 0; Gaps 0;

QY 1 AKIIDGLVAKQIREIEAIEIAEMKDAIGVPLGLAVILVGRSKDSQTYVRNKKACEAVG 60
DB 69 ATIVDGKAVAKKIRDEITIEVRMKEISGIVPLGLAVILVGRKDSQTYVRNKKACEAVG 128
QY 61 IKSYEYNLPEDSSEDEVKHIATFNSDPVSGHILVOLPLPHMNDENILNAVSIKDVGD 120
DB 129 IKSEFVRLAEDSSEVLKSVSGFNDPVSCHILVOLPLPSHMDQNLNAVSIKDVGD 188
QY 121 FHPLNIGRLAMQGRDFFVPCPTKGCMEILLHRYGVEIKGKRAVIGRSNIVGMPAALLLQ 180
DB 189 FHPLNIGRLAMGREGPLFVPCPTKGCIELLHRYNIEIKGKRAVIGRSNIVGMPAALLLQ 248
QY 181 KANATYSIVHSNTKPEEITRQADIVIAAGVANLVGRSWIKPGAAITDVGINPVDDES 240
DB 249 REDATYSIIHSRTKPEEITREADIIISAVGPNVGRSWIKPGAVLIDVGINPVDPSA 308
QY 241 PRGYRLVGVDCYEEASKIAGLTTPVPGGVPMTIAMLLSNTLESAKRIHKFK 292
DB 309 ARGYRLVGVDCYEEASKVASATTPVPGGVPMTIAMLLSNTLTSAKRIHNQ 360

RESULT 2
QBLEA3
ID Q8LEA3 PRELIMINARY; PRT; 299 AA.
AC Q8LEA3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 5,10-methylenetetrahydrofolate dehydrogenase:5,10-
DE methylenetetrahydrofolate cyclohydrolase, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

[1]
NCBI_TaxID=3702;
[2]
SEQUENCE FROM N.A.
HAAS B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Haas B.J., Flavel R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RL Genome Biol. 0:0-0(2002).
[2]
SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY085538; AAC62762.1; -.
DR GO: GO:0003924; F:catalytic activity; IEA.
DR GO: GO:0003936; P:folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR000672; THF_DHG_CyH.
DR Pfam: PF00763; THF_DHG_CyH; 1.
DR PRINTS: PR00085; THFDHGRNASE.
DR ProDom: PD002300; THF_DHG_CyH; 1.
DR PROSITE: PS00767; THF_DHG_CyH_2; 1.
SQ SEQUENCE 299 AA; 31566 MW; C934C84E91D7BAF7 CRC64;

Query Match 69.7%; Score 1035; DB 10; Length 299;
Best Local Similarity 67.4%; Pred.No. 9.5e-79;
Matches 194; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY 1 AKIIDGLVAKQIREIEAIEIAEMKDAIGVPLGLAVILVGRSKDSQTYVRNKKACEAVG 60
DB 9 AKIIDGRAIAHTIRSEIAEVRGLSEHGKVPGLAVIVGRSKDSQTYVRNKKACEAVG 68
QY 61 IKSYEYNLPEDSSEDEVKHIATFNSDPVSGHILVOLPLPHMNDENILNAVSIKDVGD 120
DB 69 IKSEFVRLAEDSSEVLKSVSGFNDPVSCHILVOLPLPSHMDQNLNAVSIKDVGD 128
QY 121 FHPLNIGRLAMQGRDFFVPCPTKGCMEILLHRYGVEIKGKRAVIGRSNIVGMPAALLLQ 180
DB 129 FHPLNIGRLAMGREGPLFVPCPTKGCIELLHRYGVEIKGKRAVIGRSNIVGMPAALLLQ 188
QY 181 KANATYSIVHSNTKPEEITRQADIVIAAGVANLVGRSWIKPGAAITDVGINPVDDES 240
DB 189 KADATYTVHSHTKPEEITREADIVIAAGQAHMKGWIKPGAAIDVGINAVSDPSK 248
QY 241 PRGYRLVGVDCYEEASKIAGLTTPVPGGVPMTIAMLLSNTLESAKRI 288
DB 249 KSGYRLVGVDCYFAASKVAGFITPVGVPMTVAMLENTVDGAKRV 296

RESULT 3
Q9ZTV0
ID Q9ZTV0 PRELIMINARY; PRT; 294 AA.
AC Q9ZTV0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE 5,10-methylenetetrahydrofolate dehydrogenase-5,10-
DE methylenetetrahydrofolate cyclohydrolase.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
CX NCBI_TaxID=3888;
[1]
SEQUENCE FROM N.A.
CHEN L., Nargang N.E., Cossins E.A.;
RT "a cDNA sequence encoding pea leaf cytosolic bifunctional 5,10-
RT methylenetetrahydrofolate dehydrogenase-5,10-methylenetetrahydrofolate
RT cyclohydrolase.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
```

RA Chen L., Nargang F.E., Cossins E.A.;
RT "Isolation and sequencing of a cDNA encoding pea cytosolic
RT functional 5,10-methylenetetrahydrofolate dehydrogenase; 5,10-
RT methylenetetrahydrofolate cyclohydrolase. Comparisons with the yeast,
RT mammalian, and bacterial proteins.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF030516; AAD01907.1; -;
DR EMBL; AJ011589; CAB56756.1; -;
DR PIR; T50664; T50664.
DR HSSP; P11586; 1A41.
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR GO; GO:0016491; F:Oxidoreductase activity; IEA.
DR GO; GO:0009396; P:Folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHGRNASE.
DR ProDom; PD02300; THF_Dhg_CYH; 1.
DR PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW Hydrolase; Oxidoreductase
SQ SEQUENCE 294 AA; 31344 MW; 5D5066EB8BA880690 CRC64;
Query Match 69.6%; Score 1034; DB 10; Length 294;
Best Local Similarity 67.2%; Pred. No. 1.1e-78;
Matches 193; Conservative 45; Mismatches 49; Indels 0; Gaps 0;
QY 1 AKIIDGLVAKQIREIAEIAIKMKDAIGVVPGLAVILVGSRKDSQSYVRNKKKACEAVG 60
DB 2 ATVIDGKAVACTIRSEIADEVRLLSQYKVPGLAVIVGNRDKDSQSYVGMKCKACALG 61
QY 61 IKSEVNLPESSDEVLKHIATNSPDSVHGILVQLPLPHNDENILNAVSEIKDVGD 120
DB 62 IKSPDIDLPEDASAEIETKIVHEINANPDVHGILVQLPLPKHNEERKVLTEISIKDVGD 121
QY 121 FHLNIGRLAMQGDPPFPCTPKGQNELHRYGVEIKGRVAVIGSNIVGMPPAALLQ 180
DB 122 FHLNIGRLAMQGRDPLFPCPKACLELLRSVSGVSKGKAVVGNRSNIVGIPASLLLL 181
QY 181 KANATSVIHSNTKKPBEITRQADIVIAVGVANLVGSMIKPGAAIIDVGINPDDPES 240
DB 182 KADATVTVHSHTSQPTIIRADIVIAAAGAKMKIGSKIPGAAIVDGTNSVDPTDPR 241
QY 241 PRGRLVGDVGYEASKIAGLITVPGVGPMTAMLLSNTLESKR 287
DB 242 KSGYRLVGDVDFEASKVAGWITVPGVGPMVTMLLNTLEGAKR 288
RESULT 4
ID Q9LHH7 PRELIMINARY; PRT; 299 AA.
AC Q9LHH7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 5,10-methylenetetrahydrofolate
DE dehydrogenase/5,10-methylenetetrahydrofolate cyclohydrolase
DE (5,10-methylenetetrahydrofolate
DE dehydrogenase; 5,10-methylenetetrahydrofolate cyclohydrolase, putative,
DE 44272-46007) (At3g12290).
GN F28115.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Farman B., Valle G., Bloeker H., Perez-Alonso M., Oberwäster B.,
RA Delseny M., Bouterne M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choisy N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmback E., Drzonek H., Exfle H., Jordan N., Bangert S., Nyakatura G.,
RA Wiedemann R., Kranz H., Voss H., Holland R., Brandt P., Simonati B.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordiek G.,
RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwälder B., Duchemin D.,
RA Cooke R., Landie M., Berger-Liauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
RA Cressy T.H., Haas B., Mailli R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Niernan W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa S., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:820-822(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
ENBL; AP02047; BAB0138.1; -;
ENBL; AC069472; AAG51064.1; -;
ENBL; BT004612; AAO42858.1; -;
HSSP; P11586; 1A41.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH_C; 1.
DR PRINTS; PR00085; THFDHGRNASE.
DR ProDom; PD02300; THF_Dhg_CYH; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ SEQUENCE 299 AA; 31589 MW; CA92CBBB3271BAF1 CRC64;
Query Match 69.3%; Score 1030; DB 10; Length 299;
Best Local Similarity 67.0%; Pred. No. 2.5e-78;
Matches 193; Conservative 46; Mismatches 49; Indels 0; Gaps 0;
QY 1 AKIIDGLVAKQIREIAEIAIKMKDAIGVVPGLAVILVGSRKDSQSYVRNKKKACEAVG 60
DB 2 ATVIDGKAVACTIRSEIADEVRLLSQYKVPGLAVIVGNRDKDSQSYVGMKCKACALG 61
QY 61 IKSEVNLPESSDEVLKHIATNSPDSVHGILVQLPLPHNDENILNAVSEIKDVGD 120
DB 62 IKSPDIDLPEDASAEIETKIVHEINANPDVHGILVQLPLPKHNEERKVLTEISIKDVGD 121
QY 121 FHLNIGRLAMQGDPPFPCTPKGQNELHRYGVEIKGRVAVIGSNIVGMPPAALLQ 180
DB 122 FHLNIGRLAMQGRDPLFPCPKACLELLRSVSGVSKGKAVVGNRSNIVGIPASLLLL 181
QY 181 KANATSVIHSNTKKPBEITRQADIVIAVGVANLVGSMIKPGAAIIDVGINPDDPES 240
DB 182 KADATVTVHSHTSQPTIIRADIVIAAAGAKMKIGSKIPGAAIVDGTNSVDPTDPR 241
QY 241 PRGRLVGDVGYEASKIAGLITVPGVGPMTAMLLSNTLESKR 287
DB 242 KSGYRLVGDVDFEASKVAGWITVPGVGPMVTMLLNTLEGAKR 288
RESULT 4
ID Q9LHH7 PRELIMINARY; PRT; 299 AA.
AC Q9LHH7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 5,10-methylenetetrahydrofolate
DE dehydrogenase/5,10-methylenetetrahydrofolate cyclohydrolase
DE (5,10-methylenetetrahydrofolate
DE dehydrogenase; 5,10-methylenetetrahydrofolate cyclohydrolase, putative,
DE 44272-46007) (At3g12290).
GN F28115.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

QY 61 IKSYENVLPDSDSEDEVKHIATFNSDPSVHGILVOLPLPHNMNDENILNAVSIKVDVG 120
DB 88 -----EDSSEEEVLYKVSFGNDPSVHGLVQLPLPSHMDQENILNAVSIKVDVG 138
QY 121 FHLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMPAALLQ 180
DB 139 FHLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMPAALLQ 198
QY 181 KANATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNVDDPES 240
DB 199 KEDATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNVDDPES 258
QY 241 PRGRLVGDVVCYEASKIAGLITPVGSGVPMTIAMLLSNTLSAKRIHKFK 292
DB 259 AGGRLVGDVVCYEASKIAGLITPVGSGVPMTIAMLLSNTLSAKRIHKFK 310
RESULT 7
Q9FRB2 PRELIMINARY; PRT; 291 AA.
AC Q9FRB2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to Pisum sativum methylenetetrahydrofolate dehydrogenase
DE (NADP+) (EC 1.5.1.5) (AF030516).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Chow T.-Y., Hsiao Y.-I.C., Chen H.-H., Wu H.-P., Chao Y.-T.,
RA Liu S.-M., Hsiao Y.-Y., Huang J.-J., Lee P.-F., Su C.-L., Chen C.-S.,
RA Shaw J.-F.;
RL "Oryza sativa PAC P0001A07 genomics sequence.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC084218; AAG48834.1; -;
DR HSSP; P11586; 1A41.
DR Gramene; Q9FRB2; -;
DR GO; GO:0003824; -:catalytic activity; IEA.
DR GO; GO:0004488; F:methylenetetrahydrofolate dehydrogenase (NA. . .; IEA.
DR GO; GO:0009386; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF00763; THF_DHG_CYH; 1.
DR Pfam; PF02882; THF_DHG_CYH; 1.
DR PRINTS; PR00085; THFDHGNASE.
DR ProDom; PD002300; THF_DHG_CYH; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
SQ SEQUENCE 291 AA; 30720 MW; 81591D06G2638C9E CRC64;
Query Match 66.8%; Score 992.5; DB 10; Length 291;
Best Local Similarity 66.6%; Pred. No. 3.4e-75;
Matches 191; Conservative 39; Mismatches 56; Indels 1; Gaps 1;
QY 1 AKIIDGKLVAKQIREIAVEIAKMDAIGVPLGLAVILVGRKSDQTVVRNKKACEAVG 60
DB 2 AQIIDGKVAADIRREVAADVAALSAHNLVPLGLAVILVGRKSDQTVVRNKKACEAVG 61
QY 61 IKSYENVLPDSDSEDEVKHIATFNSDPSVHGILVOLPLPHNMNDENILNAVSIKVDVG 120
DB 62 IRSVDVLAEDISEALVAEVRHLNADPAVHGILVOLPLPHNEEKLINLSLEKVDVG 121
QY 121 FHLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMPAALLQ 180
DB 122 FHLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMPAALLQ 181
QY 181 KANATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNVDDPES 240
DB 182 KADATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNVDDPES 240

QY 241 PRGRLVGDVVCYEASKIAGLITPVGSGVPMTIAMLLSNTLSAKR 287
DB 241 KSGYELVGDVDFAEVSKVAGHLITPVGSGVPMTIAMLLSNTLSAKR 287
RESULT 8
Q9W527 PRELIMINARY; PRT; 160 AA.
AC Q9W527;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Methylenetetrahydrofolate dehydrogenase/methylenetetrahydrofolate
DE Cylcohydrolyase isoform 2 (EC 3.5.4.9) (Fragment).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang H., Bohnert H.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF439725; AAL33591.1; -;
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004477; F:methylenetetrahydrofolate cyclohydrolyase act. . .; IEA.
DR GO; GO:0009396; P:folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_Dhg_Cyh.
DR Pfam; PF02882; THF_DHG_CYH; 1.
DR PRINTS; PR00085; THFDHGNASE.
DR ProDom; PD002300; THF_DHG_CYH; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW Hydrolase.
FT NON TER 1
FT NON TER 160 160
SQ SEQUENCE 160 AA; 16935 MW; 325D0A962AD26C2A CRC64;
Query Match 53.3%; Score 792; DB 10; Length 160;
Best Local Similarity 94.4%; Pred. No. 1.1e-58;
Matches 151; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 115 EKDVDFHPLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMP 174
DB 1 EKDVDFHPLNIGRLAMQGRDPFVCTPKGMELHRYGVEIKKRAVVGSRNIVGMP 60
QY 175 AALLLOKANATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNP 234
DB 61 AALLLOKANATVSIHVSNTKPKPEITRQADIVIAAVGVANLVRSWKFGAALIDVGNP 120
QY 235 VDDPESPRGRLVGDVVCYEASKIAGLITPVGSGVPMTI 274
DB 121 VDDPESPRGRLVGDVVCYEASKIAGLITPVGSGVPMTI 160
RESULT 9
Q9HFV0 PRELIMINARY; PRT; 937 AA.
AC Q9HFV0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Cl-THPS protein.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Langridge J., Bordoli R., Rachubinski R., Wozniak R., Smith J.,
RA Skilling J.;
RT "Direct software interpretation of de novo sequence from ESI-MS/MS
RT data for novel protein identification using BLAST.";

DR GO: GO:0003824; F: catalytic activity; IEA.
DR GO: GO:0009396; P: folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR000672; THF_DHG_CyH.
DR Pfam: PF00763; THF_DHG_CyH; 1.
DR PRINTS: PR00085; THFDHGRGNASE.
DR P: P:0002300; THF_DHG_CyH; 1.
DR PROSITE: PS00766; THF_DHG_CyH; 1.
DR PROSITE: PS00767; THF_DHG_CyH; 2; 1.
DR Complete proteome.
SQ SEQUENCE 284 AA; 30907 MW; 0413362CF693598B CRC64;

Query Match 50.8%; Score 755.5; DB 16; Length 284;
Best Local Similarity 51.68; Pred. No. 2.8e-55;
Matches 149; Conservative 60; Mismatches 71; Indels 9; Gaps 4;
QY 2 KIIDGKLVAKQIRBEETAVEIAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKACAVGI 61
DB 3 EIIDGKLVAKQIRBEETAVEIAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKACAVGI 61
QY 62 KSYEVNLPEDSSSEDEVLKHIATFNPSDPSVHGILVQLPLPHMNDENILNAVSIKDVDF 121
DB 62 KSVLIELPENVTBTEKLLSVBELNEDKTIHGILVQLPLPEHISEEKVIDTISFDKVDGF 121
QY 122 HPLNIGRLAMQGRDPFFVPCPTKGCMLLHRYGVIEIKGRAVIGRSNIVGMPAAALLQK 181
DB 122 HPNVGNLFI-GKDS-FVPCPTPAGIIEIKSTGTQIEGKRAVIGRSNIVGKPEVAQLLN 179
QY 182 ANATSVISVHNTKPEITRQADIVIAAGVANLVGRSMWIKPGAAIIVDGNPVDPEP 241
DB 180 ENATVITAHSTKDLQVAKEDILVATGLAKFVKDYIKFGAIVIDGM-----DRD 233
QY 242 RGYRLVDGVCYEEASKIAGLITPVPGGVGMTIAMLLSNTLESKRHK 290
DB 234 ENNKLCGDVDFDDVQVQAGFITPVPGGVGMTIAMLLSNTLESKRHK 282

RESULT 12

Q8WZJ7
ID Q8WZJ7 PRELIMINARY; PRT; 937 AA.
AC Q8WZJ7
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OC1-2003 (TRENBLrel. 25, Last annotation update)
DE C-1-tetrahydrofolate synthase.
GN SPC839.16.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Lyne M., Purnelle B., Goffeau A., Wood V., Rajandream M.A.,
RA Barrell B.G., Saunders D., Harris D.,
RA Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL056796; CAB46709.1; -;
DR PIR; T40723; T40723.
DR GenDB Spombe; SPC839.16; -;
DR GO: GO:0005524; F: ATP binding; IEA.
DR GO: GO:0003824; F: catalytic activity; IEA.
DR GO: GO:0004329; F: formate-tetrahydrofolate ligase activity; IEA.
DR GO: GO:0009396; P: folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR000559; Fmethyl_synch.
DR InterPro: IPR000672; THF_DHG_CyH.
DR Pfam; PF01268; FTHFS; 1.
DR Pfam; PF00763; THF_DHG_CyH; 1.
DR Pfam; PF02882; THF_DHG_CyH; 1.
DR PRINTS; PR00085; THFDHGRGNASE.
DR PROSITE; PS002300; THF_DHG_CyH; 1.
DR PROSITE; PS00721; FTHFS-1; 1.
SQ SEQUENCE 937 AA; 101202 MW; 70FF8700FD023C90 CRC64;

Query Match 50.4%; Score 749.5; DB 3; Length 937;
Best Local Similarity 50.04; Pred. No. 4.8e-54;
Matches 147; Conservative 50; Mismatches 90; Indels 7; Gaps 2;
QY 1 AKIDGKLVAKQIRBEETAVEIAKMDAIGVPGGLAVILVGRSKDSQTYVRNKKKAC 56
DB 2 ALLEGTSLARKVREELREQISSIK---SVDYFNVSLKIIQVGGREDNSVYVRNKTAA 58
QY 57 EAVGIKSYEVNLPEDSSSEDEVLKHIATFNPSDPSVHGILVQLPLPHMNDENILNAVSIK 116
DB 59 NEAGISCEHVNFPEITDYDILLAKGFNEPTVHGIIVQLPLPAHINEQIITEVAPEK 118
QY 117 DVDGPHPLNIGRLAMQGRDPFFVPCPTKGCMLLHRYGVIEIKGRAVIGRSNIVGMPAA 176
DB 119 DVDGFCETNLGKLTREGQPLFTACTPKGIMCILKHGIVQGHAVVIGRSNIVGRPMS 178
QY 177 LLLQKANATVSVHNTKPEITRQADIVIAAGVANLVGRSMWIKPGAAIIVDGNPVD 236
DB 179 ILLEKANATVILCHSKTESIADIVTADIVVAAGIPIHFVKADWLKGVKVAIDVGINSIP 238
QY 237 DPESPRGYRLVDGVCYEEASKIAGLITPVPGGVGMTIAMLLSNTLESKRHK 290
DB 239 DATKKSGLYRLTGIDIFENAKESASAITPVPGSVGMTVAMLLQNVVESAVRFRK 292

RESULT 13

Q99YX2
ID Q99YX2 PRELIMINARY; PRT; 284 AA.
AC Q99YX2
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative bifunctional methylenetetrahydrofolate dehydrogenase /
DE methylenetetrahydrofolate cyclohydrolase (EC 1.5.1.5).
GN FOLD OR SPY1502 OR SPY18 1520.
OS Streptococcus pyogenes, and
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314, 186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RA "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RA "Genome sequence and comparative microarray analysis of serotype M18
RA group A Streptococcus strains associated with acute rheumatic fever
RA outbreaks";
RA Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
RL EMBL; AE006583; AAK34300.1; -;
DR EMBL; AE010067; AAL98090.1; -;
DR HSP; P11586; 1A4I.
DR GO: GO:0004488; F: methylenetetrahydrofolate dehydrogenase (NA...); IEA.
DR GO: GO:0016491; F: oxidoreductase activity; IEA.
DR GO: GO:0009396; P: folic acid and derivative biosynthesis; IEA.
DR InterPro: IPR000672; THF_DHG_CyH.
DR Pfam; PF00763; THF_DHG_CyH; 1.
DR Pfam; PF02882; THF_DHG_CyH; 1.
DR PRINTS; PR00085; THFDHGRGNASE.
DR PROSITE; PS00085; THFDHGRGNASE.

```

DR ProDom: PD002300; THF Dhg Cyh: 1.
DR PROSITE; PS00766; THF_DHG_CYH_1; 1.
DR PROSITE; PS00767; THF_DHG_CYH_2; 1.
KW Hypothetical protein; Oxidoreductase; Complete proteome.
SQ SEQUENCE 284 AA; 31082 MW; 581EA6974ACBB35 CRC64;

Query Match 50.0%; Score 743.5; DB 16; Length 284;
Best Local Similarity 51.4%; Pred. No. 2.9e-54;
Matches 149; Conservative 55; Mismatches 77; Indels 9; Gaps 4;

QY 2 KIIDGKLVAKQIRREIAVEIAKMDAIGVPGGLAVILVGRKDSQTVYRNKKKACEAVGI 61
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 3 ELIDGKALAKMQQELAAKVNKLKQKGI VPGGLAVILVGRKDSQTVYRNKKKACEAVGI 62
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 62 KSYEVNLPEDSSEDEVLKHTATNSDPSVHGILVQLPLPHMNDENILNAVSTEKVDGDF 121
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 63 KSETVRLSEFICQELIAVIERYNADNTINGILVQLPLPHMNDENILNAVSTEKVDGDF 122
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 122 HPLNIGELAMQGRDPFVPCPKGCMELLHRYGVEIKGKAAVVGSRNIVGMPAALLQK 181
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 123 HPMTGHL-WSGR-PLMVPCTPSSIMELLREYNVNLEKHAIVIGSRNIVGMPAALLQK 180
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 182 ANATVSIHNTKPKPEITRQADIVIAAVGVANLVGRSWIKPGAAIIDVGINPVDDESP 241
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 181 KNAVTLTHSRTRQLBEVCRCAVLVAIQGHFTITKQYIKGAIIVDGMNRDNG--- 237
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 242 RGYRLVGDVCEYBASKIAGLITPVPGVGPMTIAMLLSNTLSAKR-IHK 290
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 238 ---KLIGDAFDEVAEVAAKITPVPGVGPMTIAMLLSNTLSAKR-IHK 290
Db :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 15
Q8FV26 PRELIMINARY; PRT; 299 AA.
ID Q8FV26 AC Q8FV26
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative methylenetetrahydrofolate dehydrogenase.
GN FOLD OR SPY3_1157 OR SF80705.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence".
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1 / Serotype M3;
RA Nakagawa I., Kurokawa K., Nakata M., Tomiyasu Y., Yanashita A.,
RA Yamazaki K., Okahashi N., Kawabata S., Yasunaga T., Hatcori M.,
RA Hayashi H., Hamada S.;
RT "The genome of invasive Streptococcus pyogenes; a comparative analysis
RT of S. pyogenes SSI-1, SF370 and MGAS232."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE014159; AAC79764.1; -
DR EMBL; AP005143; BAC63800.1; -
DR GO; GO:0003824; F: catalytic activity; IEA.
DR GO; GO:0009396; P: folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THF_DHG_CYH_1.
DR Pfam; PF00763; THF_DHG_CYH_1; 1.
DR PRINTS; PR00085; THFDHGRNASE.
KW Complete proteome.
SQ SEQUENCE 299 AA; 31231 MW; 7446B7B964FB967A CRC64;

Query Match 49.9%; Score 741; DB 16; Length 299;
Best Local Similarity 51.6%; Pred. No. 5e-54;

```


Matches	149;	Conservative	49;	Mismatches	89;	Indels	2;	Gaps	2;
QY	1	AKTIDGKLVAQIREIAVIAKMDAIGVVGILVGRSKDQSYVYRNKKKCAEAVG	60						
Db	2	AQLIDGKLAEDVSVTKTETELKVAATGVVGIAVIVGEDPASQYVYVASKRKACEG	61						
QY	61	IKSEYVNLNLPDSSDEVLKXIATFNSPDSVHGILVQLPLPHNMNDENILNAVIGIEKDVG	120						
Db	62	FHSVQHDLPTASEQELNLLEGLNNDPAHGLVQLPLGCHDSGRVIQTIAPEKDVG	121						
QY	121	FHPLNTGRLAMQGRDPFVFPCTPKGCMELLHR-YGVEIKKRAVVGGRSNIYGMPPAALLL	179						
Db	122	FHFINVKGLTGEVETAFVPCTPAGAIMITRVHGRDLGLNAVIGRSNIYKGMENLL	181						
QY	180	QKANATYSIVHSNTKKPEETITROADIVIAAVGANLVGRGSIWIPGAAILDVGINFPVDPE	239						
Db	182	LAANATVVAHSRTKDLPAARNADILVAAGRPQMYKGDVWEGATVIDVGINRIPAPE	241						
QY	240	SPRG-YRLVGDVCYBEASKIAGLITPVPGVGPMTIAMLLSNTLESAR	287						
Db	242	RSGSKYRLVGDVDFAEAEKVAGAITPVPGVGPMTIAMLLSNTLTAACR	290						

Search completed: July 24, 2004, 06:38:59
Job time : 87 secs

This Page Blank (uspio)